Differential Detection of Binary FM

By R. R. ANDERSON, W. R. BENNETT,
J. R. DAVEY and J. SALZ

(Manuscript received August 27, 1964)

Detection of binary FM by multiplication of the received signal by itself delayed is analyzed. Error rates vs signal-to-noise ratio for additive Gaussian noise are calculated as a function of sampling time, differential delay at the receiver, and delay distortion in the channel. It is found that the differential detector can give better performance than the more conventional zero-crossing counter or frequency discriminator under conditions of severe delay distortion in the channel.

I. INTRODUCTION

It has been found possible to realize excellent practical performance in FM transmission of binary data by use of a detector in which the signal is multiplied by a delayed replica of itself. This method has been called "differential detection" on account of its resemblance to the scheme of that name in widespread use as a detector of phase-modulated waves. The name "product demodulation" has also been applied. We can regard the detector either as a particular kind of frequency discriminator or as a phase comparator operating on the phase changes inherent in an FM wave. The former concept is suitable for a steady-state analysis, while the latter is more convenient in the study of signal transitions.

II. THE DIFFERENTIAL DELAY PRODUCTOR AS A FREQUENCY DISCRIMINATOR

Viewed as a discriminator, the detector has a steady-state response function calculable by multiplying a sine wave $A \cos (\omega_c + \omega)t$ by the corresponding delayed wave $A \cos [(\omega_c + \omega)(t - \tau)]$, i.e.,

$$A \cos (\omega_c + \omega)t \cdot A \cos [(\omega_c + \omega)(t - \tau)]$$
$$= (A^2/2) \cos (\omega_c + \omega)\tau \qquad (1)$$
$$+ (A^2/2) \cos [2(\omega_c + \omega)t - (\omega_c + \omega)\tau].$$

When the double-frequency component is suppressed by a low-pass filter, we obtain the response

$$V_{lf} = (A^2/2) \cos (\omega_c + \omega)\tau. \tag{2}$$

Consider ω as the frequency deviation in radians/sec from the midband frequency ω_c. Then a discriminator characteristic can be realized by setting $\cos \omega_c\tau = 0$, $\sin \omega_c\tau = \pm 1$, giving

$$V_{lf} = \mp(A^2/2) \sin \omega\tau. \tag{3}$$

The resulting steady-state response is nearly proportional to frequency deviation over the range in which $\sin \omega\tau$ is approximately equal to $\omega\tau$.

A linear relationship is not necessary for binary FM detection, since only the sign of the deviation is significant. Unambiguous results in the noise-free case can be secured over a range of $\omega\tau$ from $-\pi$ to π. In particular, if $\omega \doteq \pm\omega_b/2$ where ω_b is 2π times the bit rate, values of τ in the range zero to the reciprocal of the bit rate could be used, and a value equal to half the bit interval appears to be a good compromise. As τ is made small the linearity improves, and, as will be shown later, the performance approaches that of an ideal phase differentiator, which we shall refer to as a $d\varphi/dt$ detector. The latter type is of particular interest because its performance is closely approximated by either a zero-crossing counter or a frequency discriminator.

III. FM DETECTION BY DIFFERENTIAL PHASE COMPARISON

We illustrate the operation as a differential phase comparator by following a particular noise-free sequence through the detection process. The binary signal to be transmitted is shown in Fig. 1(a). It is assumed that this rectangular wave modulates the frequency of a carrier with a total shift between mark and space equal to the bit rate. This results in the phase change during a marking bit interval differing from that during a spacing bit interval by 360°. With respect to the mid-frequency as a reference, the variation of carrier phase versus time becomes $\pm 180°$ per bit interval, as indicated by the solid triangular wave of Fig. 1(b). When the channel is shaped to give a raised-cosine pulse spectrum at the demodulator input, the phase-versus-time pattern becomes rounded at the transitions approximately as shown by the dotted waveform of Fig. 1(b). The received signal is passed through a network with an envelope delay of one-half bit interval and with a phase shift at midband frequency of 270°. The phase-versus-time pattern of the delayed signal is shown by the dashed-line wave of Fig. 1(b). For a long mark interval

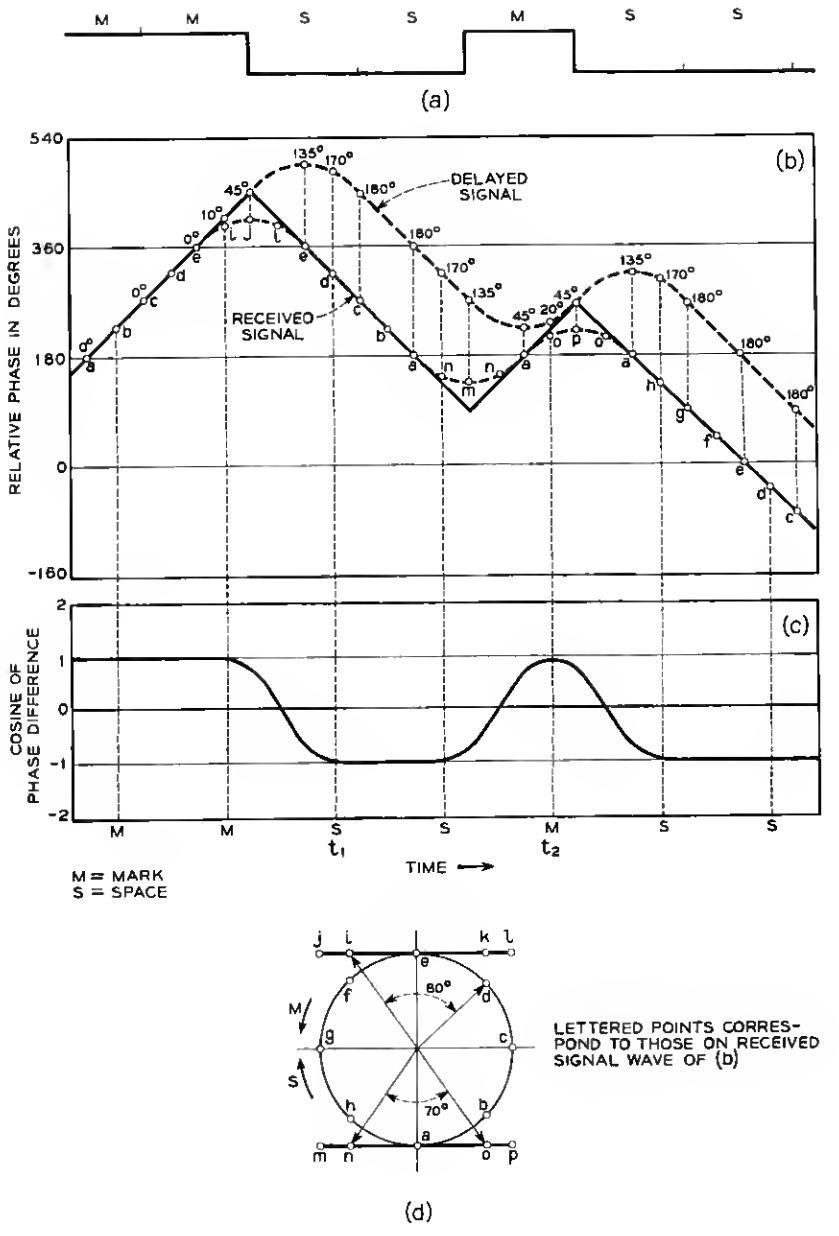

Fig. 1 — (a) Binary data, (b) phase of undelayed and delayed FM waves, (c) demodulated signal, (d) signal space diagram.

the two phase patterns are seen to he in phase, while for a long space interval they are 180° apart. The demodulation process consists of taking the product of these two waves with a switch-type modulator and filtering out thc high-frequency component. If the amplitude modulation of the received signal is negligihle, the demodulated signal becomes equal to the cosine of the phase difference hetween the delayed and undelayed versions of the signal.

The phase difference hetween the two signals at numerous points is indicated in Fig. 1(h). The shape of the demodulated wave, together with appropriate sampling times to recover the hinary information, is shown in Fig. 1(c). It will be noted that the demodulated wave is delayed one-quarter hit interval from thc instantaneous frequency of the received signal but is advanced one-quarter hit interval from the instantaneous frequency of the delayed signal.

A signal space diagram, such as described hy J. R. Davey,[1] is given in Fig. 1(d) with lettered points corresponding to the lettered points on the phase pattern of the received signal. At a transition in the demodulated signal, the vectors representing the received signal and its delayed version are in phase at points such as (i), (k), (n), and (o). At a sampling instant having a transition on only one side, such as at t_1, the vectors representing the received signal and the half-hit delayed signal are at points such as (d) and (i). These form an angle of approximately 80° which, when the delayed signal is shifted 90°, hecomes an angle of 10° or 170° depending on whether a mark or space is received. At a sampling instant having a transition on hoth sides, such as at t_2, the two vectors are at points such as (o) and (n). The angle is then about 70°, which, after the added 90° shift, results in an angle of 20° or 160°.

For a constant-amplitude signal the departures from the ideal 0° and 180° angles would decrease the amplitude of thc demodulated signal at the sampling time. The diagram of Fig. 1(d) shows, however, that at least one of the vector amplitudes at these times is ahove steady state. If the increased signal amplitude appears at the linear input of the demodulating productor, it more than compensates for the phase error, thus tending toward an overshoot of the baseband signal. The low-pass filtering will of course determine the final extent, if any, of the overshoot. It is of interest to note that a linear productor would over-emphasize these amplitude variations. Consequently, it would he expected that a switched-type modulator would result in a more perfect eye pattern.*

* By "eye pattern" is meant the oscilloscope trace obtained by sweeping the detector output against a linear time base synchronized with the bit rate. A basic description of the properties of such patterns has been given by Brand and Carter.[2]

This does not mean, however, that the probability of decision error would be different. In fact, as will be shown later, limiting one of the two inputs to the multiplier, and hence obtaining in effect a switched-type modulator, does not change the error rate.

Fig. 2 shows a computer printout of the noise-free eye pattern corresponding to one-half bit delay when the demodulator follows a product law, the total frequency shift is equal to the bit rate, and full raised-cosine spectra apply. The origin is taken at the midpoint of the bit interval in the undelayed wave. Traces of like polarity are concurrent at this point, showing the absence of intersymbol interference at these particular instants. The peak responses of the detector are not reached until a time later by half the differential delay, and the individual peaks

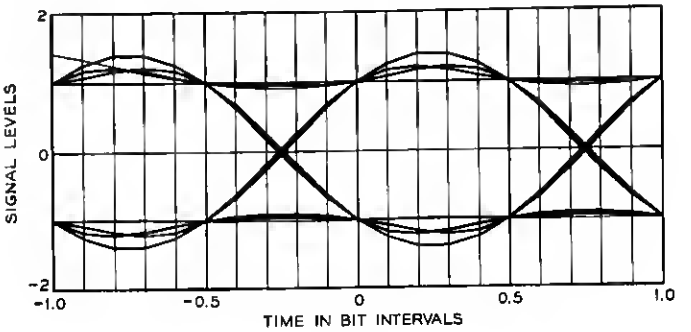

Fig. 2 — Eye pattern from product of undelayed and delayed FM waves with delay of one-half bit interval.

for different adjacent data sequences are spread over a range of values. The traces concur again at a time equal to the differential delay, and the same signal levels are observed as at the first point of concurrence. The decision-threshold (zero signal level) crossings are almost concurrent at a time preceding the origin by half the difference between the bit interval and the differential delay.

The time relations of Fig. 2 are peculiar to the choice of differential delay, ratio of frequency shift to bit rate, and spectral shaping. It will be shown later* for example, that if the differential delay is δ bit intervals, the first concurrence of traces is at the origin, the second one is at δ bit intervals later, the peaks occur at $\delta/2$, and the threshold crossings nearly at $-(1 - \delta)/2$. If the raised-cosine shaping were changed to any other satisfying Nyquist's first criterion[5] for suppression of inter-

* See Figs. 9–12 and the discussion following equation (99).

symhol interference, all concurrences would he maintained except the threshold crossings, which could spread out more. Finally, if the relation hetween frequency shift and hit rate were changed, all concurrences would ho destroyed, hut this would not hy itself imply a significant increase in error rate.

Eye patterns ohtained when a limiter is inserted in one, hut not hoth, of the two inputs are shown in Fig. 3. In Fig. 3(a), the undelayed signal is limited, giving in effect a switched modulator with the polarity of the undelayed signal switching the delayed signal. In Fig. 3(h), the delayed signal is limited, thereby interchanging the switching roles of Fig. 3(a). Comparing Fig. 3(a) with Fig. 2, we note that the concurrence of traces at 0.5 is destroyed and that the peaks of the responses are shifted to the left. In Fig. 3(h) the concurrence at 0 is destroyed and the response peaks are shifted to the right. In spite of these differences, which can he verified hy fairly straightforward analysis, the prohahility of error at a specified sampling instant must he the same for all three cases, as will be shown in detail later.

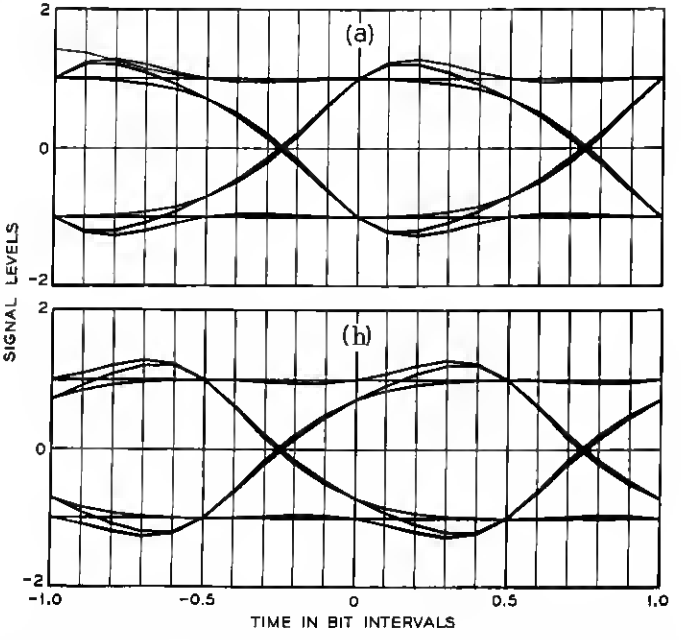

Fig. 3 — Eye patterns when one signal is limited before application to multiplier input. Delay of one-half bit interval. (a) Undelayed signal is limited; (b) delayed signal is limited.

The eye pattern for a $d\varphi/dt$ receiver, which is equivalent to the case of zero differential delay, is shown in Fig. 4. The traces concur at the peak response, which is at the origin, and the threshold crossings are half a bit interval away, as would be deduced by setting $\delta = 0$ in the discussion of the detector with differential delay. As δ approaches zero, we thus approach an equivalence with the performance of conventional frequency detectors.

The question of utility of eye patterns for nonlinear detection processes merits some further discussion. Although we cannot deduce error rates from them, we can at the least distinguish between "go" and "no go" conditions in the absence of noise. We can also use a given pattern as a basis for choosing the best sampling time. In making the choice,

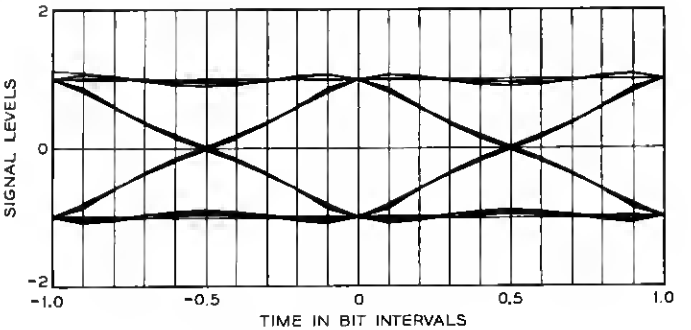

Fig. 4 — Eye pattern for $d\varphi/dt$ detector.

we consider both the horizontal and vertical margins, since relative immunity to timing jitter is important as well as the spread of sample values at the time of decision. We have verified in our calculations that the direction of change in the error rate with sampling time can be deduced by comparing eye openings for the two instants. Such deductions are not valid for comparing error rates corresponding to eye patterns for two different conditions.

Another use of the eye pattern is in laboratory diagnosis of system distortion. For such purposes we make use of established correspondence between the nature of the eye and various kinds of distortion in the particular system under test.

IV. THE EVALUATION PROBLEM

We are concerned here with the performance of the differential FM detector when the binary FM signal suffers transmission impairments.

We analyze particularly the effects of delay distortion and noise in the channel. A significant measure of performance is the curve of error rate vs signal-to-noise ratio at a specified bit rate for a channel with specified amplitude and delay variations with frequency. It is important that such a measure be determined over the range of transmission impairments encountered in actual channels. It is found that the principal virtue of the differential-delay scheme relative to the more familiar axis-crossing and frequency-discriminator types is an improved immunity to severe delay distortion.

Since there are many parameters which influence the performance of a data transmission system, the discussion would get out of hand if individual attention were given to all possible combinations of conditions. Fortunately, we can select representative regions of interest which are describable in terms of relatively few quantities. Our approach makes use of both direct analysis and digital computer back-up. In calculating the noise-free responses of the system under various conditions, we first establish formulas for a generalized data sequence. We then go to the computer to evaluate response functions vs time for a family of sequences of given length.

In theory it would be possible to calculate error rates by adding programmed noise in the computer calculation of response functions. A practical deterrent to such a procedure is the long computing time needed in the interesting cases of almost error-free transmission. It has been our experience that determination of error rates by computer simulation of additive noise is inferior to computer evaluation of analytic formulas for probability of error. Use of the latter procedure has been chiefly successful with additive Gaussian noise, but this does not necessarily imply serious limitation of utility. The premise that rank-order established on the basis of Gaussian noise holds for other kinds of interference has a good empirical foundation. A useful simplification from the Gaussian analysis is that the curves for error rate vs signal-to-noise ratio tend to be roughly parallel for different systems and to be characterized sufficiently well by their asymptotic slopes.

In terms of the normalized signal-to-noise ratio M, which is defined as the ratio of average signal power to the average white Gaussian noise power in a bandwidth equal to the bit rate, the asymptotic error rate is expressible[3] in the form $F(M) \exp(-\kappa M)$. The function $F(M)$ turns out to be of slight interest because of its relatively minor effect when M is large. For practical purposes the number κ determines the performance. For an ideal binary phase modulation system, κ has its maximum value of unity. The quantity $10 \log_{10}(1/\kappa)$ expresses noise impairment in db

relative to the ideal. It has been found possible to calculate this quantity directly without determining the entire error-rate curve.

The calculations we have made evaluate the effect of the following factors important in system design:

1. The sampling time relative to the signaling interval. The preferred sampling time is indicated by the eye patterns, but is more precisely established by error-rate calculations.

2. The length of the delay line. Equivalence of zero delay with a $d\varphi/dt$ detector establishes a reference at one end of the range in terms of a better-known system.

3. The data sequence. Results for the most and least vulnerable sequences are exhibited.

4. The delay distortion. Parabolic and linear variation with frequency are studied. The results are presented in terms of maximum delay variation expressed in bit intervals.

V. THE MODEL

A block diagram of the transmission system under study is shown in Fig. 5. The data source emits a sequence of binary symbols which for full information rate are independent of each other and have equal probability. The analysis can be generalized without analytical inconvenience by assigning a probability m_1 to one of the two binary symbols and $1 - m_1$ to the other. In conventional binary notation the symbols

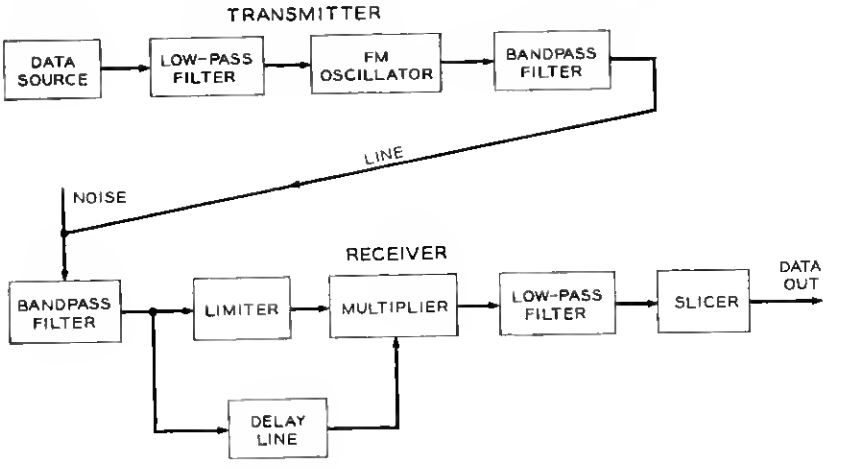

Fig. 5 — Binary FM transmission system with differential delay detection.

are 1 and 0. It is convenient to express binary frequency modulation of an oscillator in terms of positive and negative frequency deviations. The combination of data source and low-pass filter is accordingly defined by the shaped baseband data wave train

$$s(t) = \sum_{n=-\infty}^{\infty} b_n g(t - nT) \tag{4}$$

where

$$b_n = 2a_n - 1. \tag{5}$$

The values of a_n represent the data sequence in binary notation. The probability is m_1 that the typical a_n is unity, and $1 - m_1$ that it is zero. The value of b_n is $+1$ if a_n is unity, and -1 if a_n is zero. The function $g(t)$ represents a standard pulse emitted by the low-pass filter for a signal element centered at $t = 0$.

Ideally, the oscillator frequency follows the baseband signal wave $s(t)$. This would imply an output voltage from the FM oscillator specified by

$$V(t) = A \cos\left[\omega_c t + \theta_0 + \mu \int_{t_0}^{t} s(\lambda) d\lambda\right]. \tag{6}$$

Here, A is the carrier amplitude, ω_c is the frequency of the oscillator with no modulating signal applied, t_0 is an arbitrary reference time, θ_0 is the phase at $t = t_0$, and μ is a conversion factor relating frequency displacement to baseband signal voltage. The instantaneous angular frequency of the wave (6) is defined as the derivative of the argument of the cosine function. It can be written in the form $\omega_c + \omega_i$, where ω_i, the deviation from midband, is ideally expressed by

$$\omega_i = \mu s(t). \tag{7}$$

In the practical case, the transmitting bandpass filter restricts the frequency-modulated wave to the range of frequencies passed by the channel. The purpose of this filter is to prevent both waste of transmitted power in components which will not reach the receiver and contamination of the line at frequencies assigned to other channels. The result is a transformation of the voltage wave (6) to a band-limited form, which must depart in more or less degree from the ideal conditions of constant amplitude and of linear relationship between frequency and baseband signal. The line also inserts variations in amplitude- and phase-versus-frequency which cause further departures from the ideal. For our purposes it is sufficient to combine the line characteristics with those of the transmitting filter into a single composite

network function determining the wave presented to the receiving bandpass filter.

The receiving bandpass filter is necessary to exelude out-of-band noise and interference from the deteetor input. It also shapes the signal waveform and can include compensation for linear in-band distortion suffered in transmission. Two contradietory attributes are sought in the filter — a narrow band to reject noise and a wide band to supply a good signal wave to the detector. Previous work[3] has indicated a eosine filter as a near optimum.

The noise-free input to the deteetor will be written in the form

$$V_r(t) = P(t) \cos(\omega_c t + \theta) - Q(t) \sin(\omega_c t + \theta). \tag{8}$$

$P(t)$ and $Q(t)$ represent in-phase and quadrature signal modulation components respectively, whieh are assoeiated with a earrier wave at the midband frequeney ω_c with speeified phase θ. Sueh a resolution can always be made, even though the details in actual examples may be burdensome. The added noise wave at the deteetor input is assumed to be Gaussian with zero mean and ean likewise be written as

$$v(t) = x(t) \cos(\omega_c t + \theta) - y(t) \sin(\omega_c t + \theta). \tag{9}$$

If $v(t)$ represents Gaussian noise band-limited to $\pm 2\omega_c$, $x(t)$ and $y(t)$ are also Gaussian and are band-limited to $\pm\omega_c$. If the speetral density of $v(t)$ is $w_v(\omega)$, the spectral densities of $x(t)$ and $y(t)$ are given by[3]

$$w_x(\omega) = w_y(\omega) = w_v(\omega_c + \omega) + w_v(\omega_c - \omega), \qquad |\omega| < \omega_c. \tag{10}$$

In general, $x(t)$ and $y(t)$ are dependent, with eross-spectral density

$$w_{xy}(\omega) = j[w_v(\omega_c - \omega) - w_v(\omega_c + \omega)] \tag{11}$$

and cross-eorrelation functiou expressed in terms of $R_v(\tau)$, the auto-eorrelation funetion of $v(t)$, by

$$R_{xy}(\tau) = -2R_v(\tau) \sin \omega_c \tau. \tag{12}$$

The eross-correlation vanishes at $\tau = 0$, and hence the joint distribution of $x(t)$, $y(t)$ at any speeified t is that of two independent Gaussian variables.

A conveuient analytieal model of the detector is a multiplier with delayed and undelayed waves applied as inputs and with a low-pass filter in the output to seleet the difference-frequeney components of the produet. In praetieal systems various departures from the basie model may offer a more convenient realization by physical eireuits. The pure product law can be approximated by a switched modulator, an example of which is shown in Fig. 6(a). Here one of the two inputs operates a

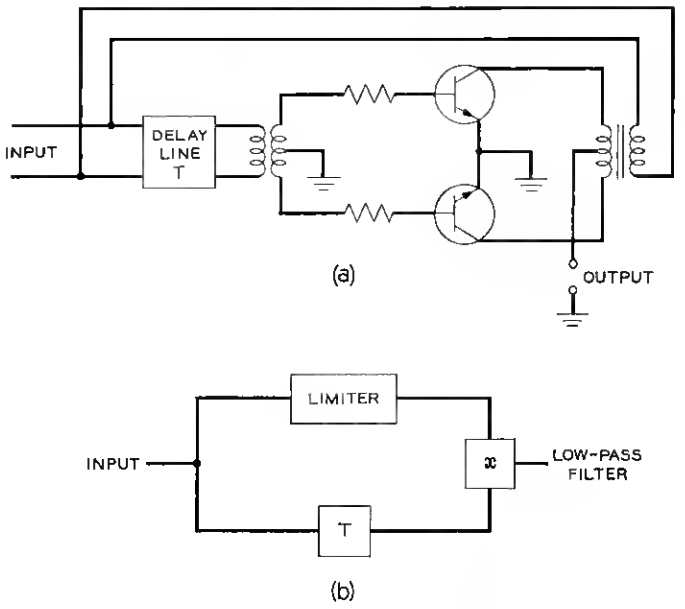

Fig. 6 — Differential detector with switched modulator.

two-transistor reversing switch in the path of the other input. The second input wave is in effect multiplied by a rectangular wave with axis crossings determined by the first input. The same result would be obtained from a strictly analog multiplier if the amplitude of one input were severely limited beforehand as shown in Fig. 6(b). We will base our analysis on a true productor with no limiter, but we show in Appendix A that limiting one input does not affect the results in the narrow-band case. In this detector the noise waves associated with the delayed and undelayed signal inputs are correlated. The amount of correlation depends on the value of the delay. The value of the delay relative to one bit interval will remain as a parameter to be optimized. However, we will require that the delay line should be designed to have a phase shift equal to an odd multiple of $\pi/2$ radians at the midband angular frequency ω_c.

VI. ANALYTICAL SOLUTION

By adding (8) and (9) we obtain the signal at the input of the detector

$$E(t) = x_1(t) \cos \omega_c t - y_1(t) \sin \omega_c t \tag{13}$$

where

$$\begin{aligned} x_1(t) &= x(t) + P(t) \\ y_1(t) &= y(t) + Q(t). \end{aligned} \tag{14}$$

The delayed signal is,

$$E(t-\tau) = x_1(t-\tau)\cos(\omega_c t - \omega_c\tau) - y_1(t-\tau)\sin(\omega_c t - \omega_c\tau) \tag{15}$$

where

$$0 < \tau < T.$$

In the case in which $\omega_c\tau = 270°$, we write for $E(t-\tau)$

$$E_d(t) = -x_{1d}\sin\omega_c t - y_{1d}\cos\omega_c t \tag{16}$$

where we set $f(t-\tau) \triangleq f_d$.

The low-frequency component $E_{lf}(t)$ of the product $E(t)E_d(t)$ is given by

$$2E_{lf} = y_1x_{1d} - x_1y_{1d} = \xi. \tag{17}$$

The performance of our system can be studied by analyzing the probability distribution of the quadratic form in (17). Since this is a binary system we only need the distribution at one point, namely "zero."

By a relabeling of the variables in which y_1, x_1, y_{1d}, and x_{1d} are replaced by x_1, x_2, x_3, and x_4, respectively, the calculation is reduced to the single problem of evaluating the probability that the quadratic form $x_1x_4 + x_2x_3$ is negative or positive, where x_1, x_2, x_3 and x_4 are Gaussian random variables with equal variances σ^2 and mean values given by:

$$\begin{aligned} E\{x_1\} &= Q(t), & E\{x_2\} &= -P(t) \\ E\{x_3\} &= Q(t-\tau), & E\{x_4\} &= P(t-\tau). \end{aligned} \tag{18}$$

We remark that all these average values, in general, will depend on the signal sequence.

A solution of this general problem in terms of an integral has been found[3] when the variables are independent. The present case is more complicated in that with an arbitrary delay τ, the noise samples become dependent. We must now include a nonzero covariance of x_1 and x_{1d} and of y_1 and y_{1d}. We point out that a solution based on uncorrelated noise samples would indicate a considerably poorer performance than found when the actual correlation is included. This is a case in which noise correlation is beneficial rather than harmful.

It has been found possible to apply the previous solution for the independent case by subjecting the four dependent Gaussian variables to a linear transformation. The new set of variables z_1, z_2, z_3, and z_4 becomes independent while preserving the invariance of the quadratic form on which decisions are based, i.e.,

$$z_1z_4 + z_2z_3 = x_1x_4 + x_2x_3 . \tag{19}$$

The nonzero covariances of the variables are:

$$\begin{aligned} \text{cov } (x_1 , x_3) &= \text{cov } (y,y_d) = r\sigma^2 \\ \text{cov } (x_2 , x_4) &= -\text{cov}(x,x_d) = -r\sigma^2. \end{aligned} \tag{20}$$

The value of r is the normalized autocorrelation function of the noise evaluated at lag time τ. The autocorrelation function is the Fourier transform of the spectral density. In the optimum receiver design the filter preceding the detector has a cosine amplitude-frequency response reaching zero at a frequency displacement from midband equal to the bit rate. Hence with white Gaussian noise on the line, the spectral density of the noise at the detector input is a squared- or raised-cosine function. Such a spectral function has the property that its Fourier transform, the autocorrelation function, decreases to zero when the lag time increases from zero to T. The solution we shall give is valid for any value of r within the permissible range from -1 to $+1$.

The transformation that satisfies (19) and at the same time diagonalizes the covariance matrix of the z variables is:

$$\begin{aligned} z_1 &= \tfrac{1}{2}(x_1 - x_2 + x_3 + x_4) \\ z_2 &= \tfrac{1}{2}(x_1 + x_2 + x_3 - x_4) \\ z_3 &= \tfrac{1}{2}(-x_1 + x_2 + x_3 + x_4) \\ z_4 &= \tfrac{1}{2}(x_1 + x_2 - x_3 + x_4). \end{aligned} \tag{21}$$

It can be verified that (19) is satisfied when this transformation is applied and also that if the correlation matrix of the x's is:

$$R_x = \sigma^2 \begin{bmatrix} 1 & 0 & r & 0 \\ 0 & 1 & 0 & -r \\ r & 0 & 1 & 0 \\ 0 & -r & 0 & 1 \end{bmatrix} \tag{22}$$

the correlation matrix of the z's becomes:

$$R_z = \sigma^2 \begin{bmatrix} 1+r & 0 & 0 & 0 \\ 0 & 1+r & 0 & 0 \\ 0 & 0 & 1-r & 0 \\ 0 & 0 & 0 & 1-r \end{bmatrix}. \tag{23}$$

We now recall the previously obtained result[3] for the four uncorrelated Gaussian variables z_1, z_2, z_3, and z_4, with the pair z_1 and z_2 having equal variances σ_0^2 and the pair z_3 and z_4 having equal variances σ_1^2:

$$\text{Prob}\,[(z_1z_4 + z_2z_3) < 0 \quad \text{when} \quad \bar{z}_1\bar{z}_4 + \bar{z}_2\bar{z}_3 > 0] = \Lambda(\rho,a,b)$$
$$= \frac{1}{2} - \frac{\rho}{2\sqrt{\pi}} \int_{-1}^{1} \exp(-\rho^2x^2)\ \text{erf}\,[a\rho(1-x^2)^{\frac{1}{2}} - bx]dx \tag{24}$$

where

$$\rho^2 = \frac{\bar{z}_1^2 + \bar{z}_2^2}{2\sigma_0^2} \tag{25}$$

$$a = \frac{\bar{z}_1\bar{z}_4 + \bar{z}_2\bar{z}_3}{2\sigma_1\sigma_2\rho^2} \tag{26}$$

$$b = \frac{\bar{z}_2\bar{z}_4 - \bar{z}_1\bar{z}_3}{2\sigma_1\sigma_2\rho}. \tag{27}$$

In our case

$$\sigma_0^2 = (1+r)\sigma^2, \qquad \sigma_1^2 = (1-r)\sigma^2. \tag{28}$$

It is shown in Appendix B that the integral in (24) can be simplified and reduced to a two-parameter form. We also show how the general asymptotic expression for large signal-to-noise ratio is derived.

The simplified version of (24) is

$$\Lambda(\rho,a,b) = \frac{1}{2\pi} \int_0^{\pi} \exp\left[\frac{-c^2}{1 + d^2\cos\varphi}\right] d\varphi \tag{29}$$

where

$$c^2 = \frac{2a^2\rho^2}{k^2 + a^2 + 1} \tag{30}$$

$$d^2 = \frac{[(k^2 + a^2 - 1)^2 + 4k^2]^{\frac{1}{2}}}{k^2 + a^2 + 1} \tag{31}$$

and $k = b/\rho$.

The signal-to-noise ratio appears only in the parameter ρ. The asymptotic expression for large signal-to-noise ratio is therefore obtained as the limiting form of (29) when c^2 becomes large. The result, as shown in Appendix B, is:

$$\Lambda(\rho,a,b) \equiv G(c,d) \sim \frac{1+d^2}{2cd\sqrt{2\pi}} \exp\left[\frac{-c^2}{1+d^2}\right]. \tag{32}$$

It is of particular interest in our problem to write the asymptotic expression in the form:

$$G(c,d) = \frac{\alpha}{\rho} e^{-\beta\rho^2} \tag{33}$$

The values of α and β are:

$$\begin{aligned} \alpha &= \frac{k^2+a^2+1+[(k^2+a^2-1)^2+4k^2]^{\frac{1}{2}}}{4\pi^{\frac{1}{2}}a[(k^2+a^2-1)^2+4k^2]^{\frac{1}{4}}} \\ \beta &= \frac{2a^2}{k^2+a^2+1+[(k^2+a^2-1)^2+4k^2]^{\frac{1}{2}}}. \end{aligned} \tag{34}$$

These are the general results we need.

VII. A REPRESENTATIVE EXAMPLE

A convenient analytical representation of a band-limited FM signal was first proposed by Sunde.[4] This representation approximates very closely the actual signals generated by practical data sets.[3] In the case of ideal transmission, we write for the signal at the input to the detector

$$E(t) = A \sin \omega_d t \sin \omega_c t - A s_1(t) \cos \omega_c t \tag{35}$$

where

$$s_1(t) = \sum_{n=-\infty}^{\infty} (-1)^n b_n g(t - nT) \tag{36}$$

$$T\omega_d = \pi, \qquad T \triangleq \text{bit interval.}$$

The standard pulse response $g(t)$ must satisfy Nyquist's first criterion, i.e.,

$$g(mT) = \delta_{m0}\,. \tag{37}$$

The value of b_n is $+1$ for mark and -1 for space.

As can be seen, the signal (35) can be synthesized by exciting a network having impulse response $g(t)$ by a series of $+$ or $-$ impulses occurring at integral multiples of the bit interval T. The sinusoidal com-

ponents of the signal are added to the response to the impulses. When delay distortion is present, the impulse response $g(t)$ is suitably modified by introducing a quadrature component.

The delayed version of (35) becomes:

$$E_d(t) = A \sin(\omega_d t - \omega_d T\delta) \sin(\omega_c t - \omega_c T\delta) - As_1(t - T\delta) \cos(\omega_c t - \omega_c T\delta) \tag{38}$$

where $\delta = \tau/T$. When the noise components are added and the condition $\omega_c T\delta = 3\pi/2$ is satisfied, the low-frequency product is

$$\xi(t) = -[A \sin(\omega_d t - \pi\delta) + y(t - T\delta)][As_1(t) + x(t)] + [A \sin \omega_d t + y(t)][As_1(t - T\delta) + x(t - T\delta)]. \tag{39}$$

The function $\xi(t)$ represents the signal as it appears at the output of the ideal low-pass filter. To obtain the information, $\xi(t)$ must be sampled at integral multiples of the bit interval. The sample thus obtained at $t = T + \epsilon T$, $0 \leqq \epsilon \leqq 1$ is

$$\begin{aligned}\xi(T + T\epsilon) = &-\{A \sin[\pi(1 + \epsilon - \delta)] + y[T(1 + \epsilon - \delta)]\} \\ &\cdot\{As_1[T(1 + \epsilon)] + x[T(1 + \epsilon)]\} \\ &+ \{A \sin[\pi(1 + \epsilon)] + y[T(1 + \epsilon)]\} \\ &\cdot\{As_1[T(1 + \epsilon - \delta)] + x[T(1 + \epsilon - \delta)]\}.\end{aligned} \tag{40}$$

As is characteristic of nonlinear detection processes, the presence of noise introduces dependence on signal history. The memory can be minimized by using a pulse spectrum which satisfies the second as well as the first of Nyquist's criteria,[5] i.e., one which preserves the spacing of transition times as well as axis crossings. To illustrate let $\epsilon = 0$ and $\delta = \frac{1}{2}$. In this case (40) reduces to:

$$\xi(T) = -\{A + y(T/2)\}\{s_1(T) + x(T)\} + \{y(T)\}\{As_1(T/2) + x(T/2)\} \tag{41}$$

where

$$s_1(t) = \sum_{n=-\infty}^{\infty} (-1)^n b_n g(t - nT) \tag{42}$$

$$s_1(T/2) = -\tfrac{1}{2}(b_1 - b_0), \qquad s_1(T) = -b_1. \tag{43}$$

The memory is thus reduced to one previous symbol. It follows, therefore, that

$$\xi(T) = [x - Ab_1][-y_d - A] + y[x_d - (A/2)(b_1 - b_0)]. \tag{44}$$

The probability of error is the weigbted average of the eonditional probabilities tbat the sample is negative when b_1 is $+1$ and that the sample is positive when $b_1 = -1$. Sinee the sample depends on the present and immediately preceding symbols, tbere are four different eases to consider: $b_1 = -b_0 = 1$; $b_1 = b_0 = 1$; $b_1 = -b_0 = -1$; and $b_1 = b_0 = -1$. If marking and spacing symbols are equally probable, eaeh ease has a probability of one-fourth. We now write for the average probahility of error

$$P_e = \tfrac{1}{4} \sum_{n=1}^{4} \text{Prob } [(x_{1n}x_{4n} + x_{2n}x_{3n}) < 0] \tag{45}$$

where the variables x_{1n} , x_{2n} , x_{3n} , and x_{4n} are speeified by Table I.

A physical interpretation of the detection process in this case can be obtained by regarding the quadratie form in (44) as the scalar product of two veetors, e.g.:

$$x_1x_4 + x_2x_3 = \mathbf{u} \cdot \mathbf{v} \tag{46}$$

where if **i**,**j** represent unit veetors along rectangular eoordinate axes, the four possible pairs of vectors are:

$$\begin{aligned} \mathbf{u}_1 &= \mathbf{i}(-A + x) + \mathbf{j}y \\ \mathbf{v}_1 &= \mathbf{i}(-A - y_d) + \mathbf{j}(-A + x_d) \end{aligned} \tag{47}$$

$$\begin{aligned} \mathbf{u}_2 &= \mathbf{i}(-A + x) + \mathbf{j}y \\ \mathbf{v}_2 &= \mathbf{i}(-A - y_d) + \mathbf{j}x_d \end{aligned} \tag{48}$$

$$\begin{aligned} \mathbf{u}_3 &= \mathbf{i}(A + x) - \mathbf{j}y \\ \mathbf{v}_3 &= \mathbf{i}(A + y_d) + \mathbf{j}(A + x_d) \end{aligned} \tag{49}$$

$$\begin{aligned} \mathbf{u}_4 &= \mathbf{i}(A + x) - \mathbf{j}y \\ \mathbf{v}_4 &= \mathbf{i}(A + y_d) + \mathbf{j}x_d \, . \end{aligned} \tag{50}$$

Oeeurrenee of error is synonymous with a negative value for the sealar produet of any pair of vectors, and henee is also equivalent to an

TABLE I

n	x_{1n}	x_{2n}	x_{3n}	x_{4n}
1	$x - A$	y	$x_d - A$	$-y_d - A$
2	$x - A$	y	x_d	$-y_d - A$
3	$x + A$	$-y$	$x_d + A$	$y_d + A$
4	$x + A$	$-y$	x_d	$y_d + A$

angle greater than 90° between the vectors of a pair. We can use this representation to explain the beneficial effects of correlated noise samples. There is positive correlation between x and x_d as well as between y and y_d. This means that x_d is likely to have the same sign as x and similarly for y_d and y. Inspection of the vector components shows that such noise values have opposite effects on the size of the angle between the two vectors, as shown in Fig. 7, illustrating (49).

At this point we can obtain an explicit expression for the probability of error. From Table I, we note that the average values of the x's are given by Table II. From (21), the $\bar{z}$'s are given by Table III. It is seen that the cases of $n = 3$ and 4 differ from $n = 1$ and 2 respectively by a change of sign throughout. A change of sign in all the $\bar{z}$'s in (25)–(27)

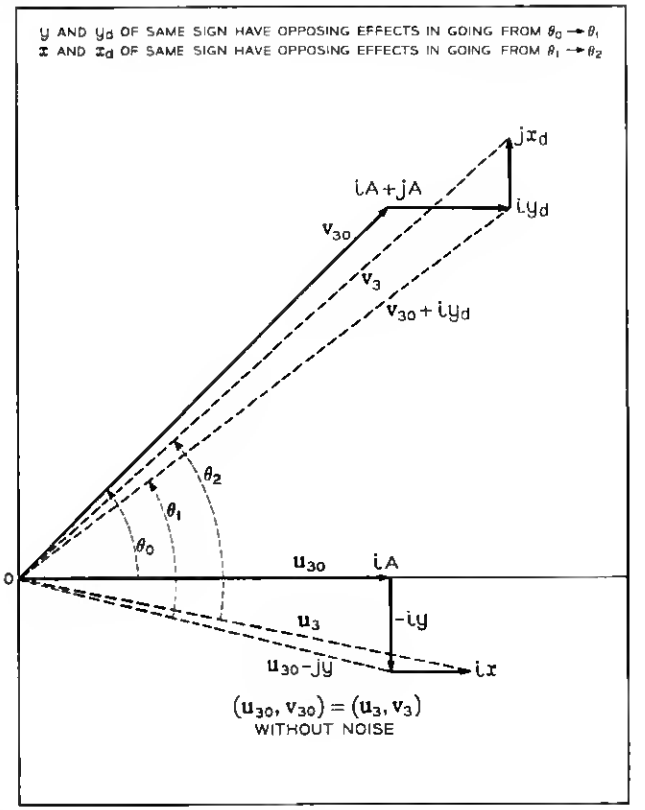

Fig. 7 — Effect of correlated noise samples on angle between vector inputs to detector.

TABLE II

n	$\bar{x}_{1n}$	$\bar{x}_{2n}$	$\bar{x}_{3n}$	$\bar{x}_{4n}$
1	$-A$	0	$-A$	$-A$
2	$-A$	0	0	$-A$
3	A	0	A	A
4	A	0	0	A

leaves the parameters a and b unchanged. Therefore, the four probabilities of (45) reduce to two distinct ones with the multiplying factor changed from $\frac{1}{4}$ to $\frac{1}{2}$.

The values of the parameters for the two cases are found to be

$$\rho_1^2 = \frac{5A^2}{4(1+r)\sigma^2} \qquad \rho_2^2 = \frac{A^2}{2(1+r)\sigma^2}$$

$$a_1 = \frac{2}{5}\sqrt{\frac{1+r}{1-r}} \qquad a_2 = \sqrt{\frac{1+r}{1-r}} \tag{51}$$

$$b_1 = -\frac{A}{2\sigma\sqrt{5(1-r)}} \qquad b_2 = 0.$$

As demonstrated in Ref. 3, the mean square of the signal wave on the line with random data is A^2 in the optimum case for white Gaussian noise. The noise input to the detector has a squared-cosine spectral density function, with σ^2 equal to the area. Hence σ^2 is one-half the mean-square value of the noise on the line in a band of width equal to twice the bit rate. In terms of the parameter M, defined as the ratio of average signal power on the line to average noise power in the bit-rate bandwidth, we have

$$M = A^2/\sigma^2. \tag{52}$$

Substituting the appropriate correlation value of $r = \frac{1}{2}$, we obtain

TABLE III

n	$\bar{z}_{1n}$	$\bar{z}_{2n}$	$\bar{z}_{3n}$	$\bar{z}_{4n}$
1	$-3A/2$	$-A/2$	$-A/2$	$-A/2$
2	$-A$	0	0	$-A$
3	$3A/2$	$A/2$	$A/2$	$A/2$
4	A	0	0	A

for the two sets of parameters

$$\begin{aligned} \rho_1^2 &= \frac{5}{6}M & \rho_2^2 &= \frac{M}{3} \\ a_1 &= \frac{2}{5}\sqrt{3} & a_2 &= \sqrt{3} \\ b_1 &= -\sqrt{\frac{M}{10}} & b_2 &= 0. \end{aligned} \tag{53}$$

The probability of error is given by

$$P_e = \tfrac{1}{2}\,\Lambda(\rho_1, a_1, b_1) + \tfrac{1}{2}\,\Lambda(\rho_2, a_2, b_2). \tag{54}$$

Using the computer program previously established for $\Lambda(\rho,a,b)$, S. Habib has obtained the uppermost curve of Fig. 8. Also shown are the ideal performance for coherent binary detection, the ideal performance for noncoherent binary detection, and an experimental curve obtained by E. R. Day.

It is also instructive to apply the asymptotic formulas given in (33). In this example, applying (53) to (34), we obtain

$$\begin{aligned} \alpha_1 &= \frac{3}{4}\sqrt{\frac{5}{3\pi}}, & \beta_1 &= \frac{2}{5} \\ \alpha_2 &= \frac{1}{2}\sqrt{\frac{3}{2\pi}}, & \beta_2 &= 1. \end{aligned} \tag{55}$$

Substituting these values in (30) we find

$$P_e \sim \frac{3}{4\sqrt{2\pi M}}\,e^{-M/3} + \frac{3}{8\sqrt{2\pi M}}\,e^{-M/3} \sim \frac{9}{8\sqrt{2\pi M}}\,e^{-M/3}. \tag{56}$$

In terms of the notation introduced in Section IV, $\kappa = \frac{1}{3}$. The asymptote is plotted in Fig. 8 and is found to be very close to the result of the exact calculation in the region of interest. Since the optimum coherent system has an error probability proportional to e^{-M}, the half-bit differential system in the limit requires $10 \log_{10} 3$ or 4.8 db more signal-to-noise ratio than the optimum for the same performance. Of this penalty, 3 db is accounted for by the steady-state informationless tones which make up half the average power of the FM1 signal. The remaining 1.8 db can be ascribed to the differential-detection scheme.

It is also possible to decode the message by providing a full-bit delay at the receiver. However, the decoding has to be performed on the transitions. In addition, the phase shift must be such that $\cos \omega_c T = 1$.

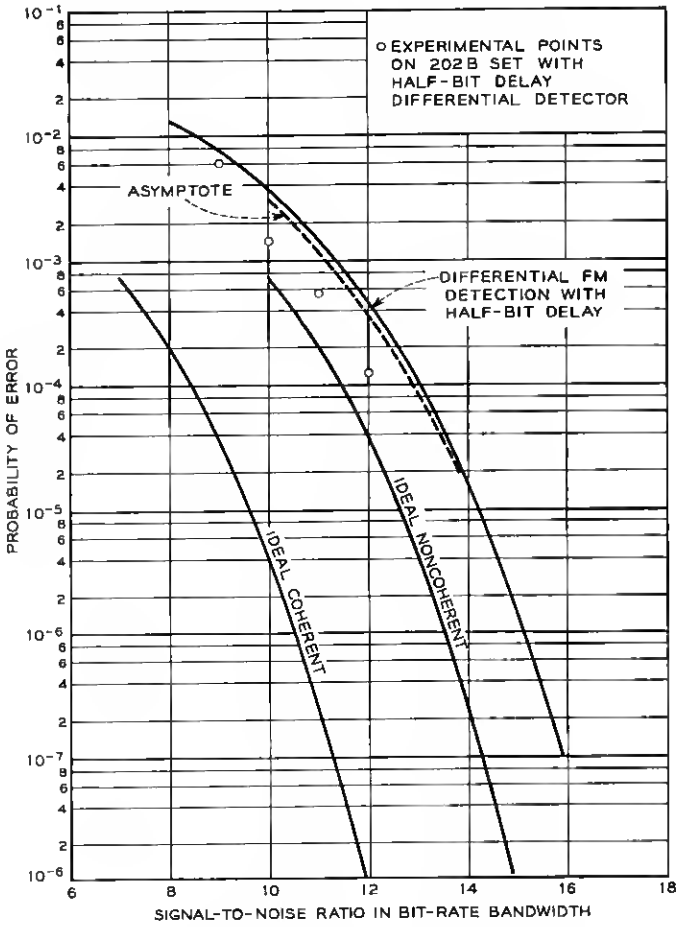

Fig. 8 — Error rates in binary data transmission.

As before, if we perform the indicated multiplication, reduce to a sum of single-frequency terms, and reject those that contain the midband frequency ω_c, we obtain the post-detection filter output, which we now designate as $V_{1f}(t)$. Again, to simplify the formulas we omit showing the functional dependence on t and use the subscript d to indicate values of functions at $t - T$. Then

$$2V_{1f} = x_1x_{1d} - x_1As_{1d} - x_{1d}As_1 + A^2s_1s_{1d} + A(y_1 - y_{1d}) \sin \omega_d t + y_1y_{1d} - A^2 \sin^2 \omega_d t. \tag{57}$$

Finally, we assume samples to be taken at multiples of T, which means that sin $\omega_d t$ becomes sin $\omega_d mT = 0$ and

$$s_1 = s_1(mT) = (-)^m b_m \tag{58}$$

$$s_{1d} - s_1[(m-1)T] = (-)^{m-1} b_{m-1} . \tag{59}$$

For the mth sample, then

$$2V_{1f} = x_1 x_{1d} + y_1 y_{1d} - (-)^m A(b_m x_{1d} - b_{m-1} x_1) - A^2 b_m b_{m-1} . \tag{60}$$

Let the binary information be coded in the transitions. Then, for example, a "1" is represented by $b_m = -b_{m-1}$ and a "0" by $b_m = b_{m-1}$. The two signaling conditions are then:

$$\begin{aligned} 2V_{1f1} &= A^2 + (-)^m b_m A(x_1 + x_{1d}) + x_1 x_{1d} + y_1 y_{1d} \\ &= [A + (-)^m b_m x_1][A + (-)^m b_m x_{1d}] + y_1 y_{1d} \end{aligned} \tag{61}$$

$$\begin{aligned} 2V_{1f0} &= -A^2 + (-)^m b_m A(x_1 - x_{1d}) + x_1 x_{1d} + y_1 y_{1d} \\ &= -[A - (-)^m b_m x_1][A + (-)^m b_m x_{1d}] + y_1 y_{1d} . \end{aligned} \tag{62}$$

The sampling interval T is typically large enough in this case to make the delayed noise samples independent of the direct samples. With this assumption we can regard x_1, y_1, x_{1d}, and y_{1d} as independent Gaussian variables. In the first condition, we set

$$\xi = A + (-)^m b_m x_1 \tag{63}$$

$$\xi_d = A + (-)^m b_m x_{1d} \tag{64}$$

$$2V_{1f} = \xi\xi_d + y_1 y_{1d} . \tag{65}$$

Then ξ, ξ_d, y_1, and y_{1d} are independent Gaussian variables with standard deviation σ, where σ is the rms noise voltage at the detector input, i.e., the rms value of either x_1 or y_1. The mean values of y_1 and y_{1d} are zero, and the mean values of ξ and ξ_d are A. In the second condition, we set

$$\xi = -A + (-)^m b_m x_1 \tag{66}$$

and obtain the same relations except that the mean value of ξ becomes $-A$ instead of A.

Correct decisions are made in the first condition if $2V_{1f}$ is positive, and in the second condition if $2V_{1f}$ is negative. Hence if we let

$$z = \xi\xi_d + y_1 y_{1d} \tag{67}$$

and designate $p_1(z)$ and $p_2(z)$ as the probability density functions of z

when ξ has the means A and $-A$ respectively, the probabilities of error in the two conditions become

$$P_1 = \int_{-\infty}^{0} p_1(z)dz \tag{68}$$

$$P_2 = \int_{0}^{\infty} p_2(z)dz. \tag{69}$$

This is the same problem solved in equations (54)–(58) of a previous paper[3], and the final result is found to be

$$P = P_1 = P_2 = \tfrac{1}{2}\, e^{-A^2/(2\sigma^2)} = \tfrac{1}{2}\, e^{-M/2}. \tag{70}$$

Equation (70) shows 3 db poorer performance than ordinary differentially coherent phase modulation. The one-bit delay differential system with transition coding thus suffers 3.9 db penalty relative to ideal coherent detection in the error probability range of 10^{-4}.

An interesting result is obtained when the delay line at the receiver is allowed to become small relative to the bit interval while still maintaining the condition that the phase shift at the carrier frequency ω_c equals 270°. We show in Appendix C that the performance in this case approaches the performance of an ideal phase differentiator.

VIII. RESULTS WITH NO DELAY DISTORTION

The first part of our numerical results deals with the performance of the differential FM detector as a function of the value of differential delay and the sampling instant in the absence of delay distortion on the line. As a preliminary, we show in Figs. 9–12 inclusive the computer print-outs of the eye patterns for differential-delay values of 0.2, 0.4,

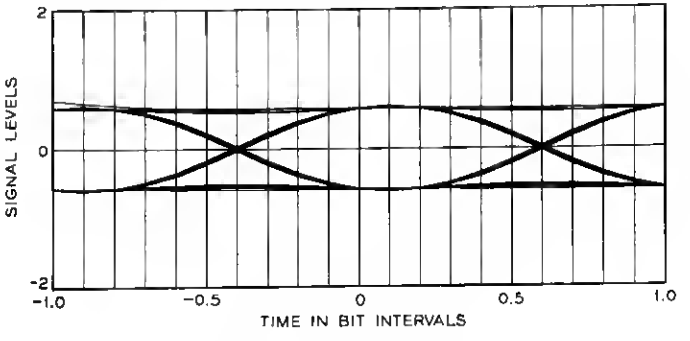

Fig. 9 — Eye pattern for differential delay of 0.2-bit interval.

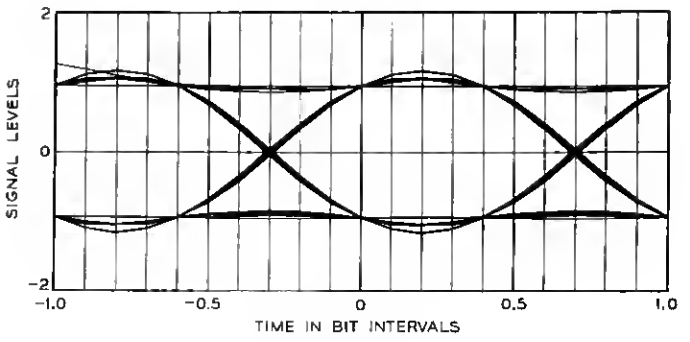

Fig. 10 — Eye pattern for differential delay of 0.4-bit interval.

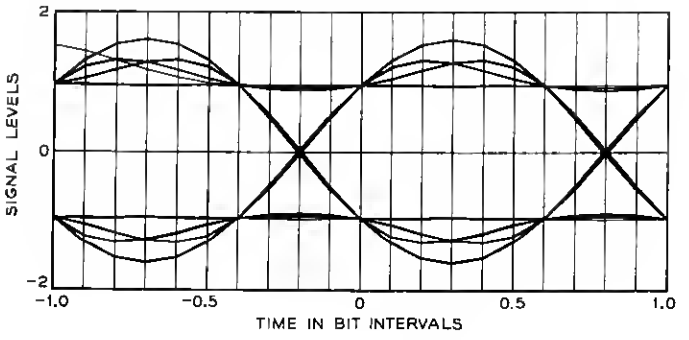

Fig. 11 — Eye pattern for differential delay of 0.6-bit interval.

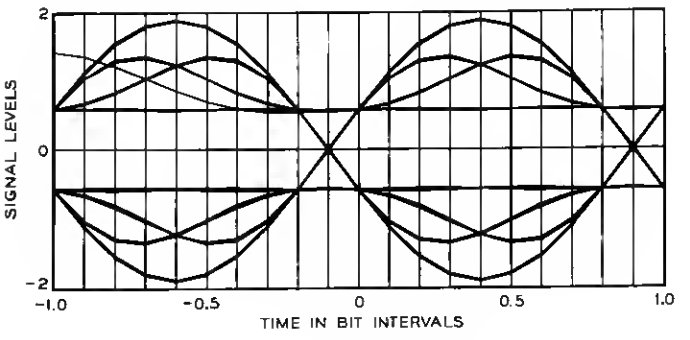

Fig. 12 — Eye pattern for differential delay of 0.8-bit interval.

0.6, and 0.8 bit interval. Some of the features found bave been previously mentioned in Sec. III. As before stated, tbe patterns are not to be interpreted as furnishing quantitative evaluation of performance. For the latter, we rely entirely on error rates vs signal-to-noise ratio. The figures illustrate wbat would be seen at tbc detector output in tbe absence of noise for the various cases and indicate tbe preferred sampling time for eacb case.

For any pair of values of δ and ϵ in (40), ξ is a random variable possessing a probability density determined by tbe additive noise and previous signal history. We have conserved computation time without loss of essential information by concentrating attention on tbe asymptotic performance when the signal-to-noise ratio is large. The corresponding db impairment or degradation relative to optimum binary PM as expressed by $10 \log_{10} (1/\kappa)$ has been computed for various values of δ and ϵ and for all data sequences of length 5 bits. For a few representative cases, we have used the exact formula over a range of signal-to-noise ratios to indicate thc degrec of approximation given by the asymptotic formula for typical conditions of intcrest. A raised-cosine pulse spectrum on the line has been assumed. Details of the computations are given in Appendix D.

The IBM 7090 computer was programmed to evaluate the parameters developed in Appcndix D for all 32 possible 5-bit sequences and for different values of δ and ϵ. Figs. 13–18 represent tbe degradation vs value of delay for different sampling instants across thc bit. The sampling time is measured from the midpoint of the bit interval in the undelayed wave. The curves on each graph represent the various sets of sequences. It appears from the graphs that the 32 sequences tend to bunch into four distinct sets. We did not attempt to label or identify tbese sets, since the average performance is more closely determined by the sequences which suffer the most. This is because the relative probability of error for the sets varies exponentially with the degradation. It can be seen that as the valuc of delay exceeds half a bit the performance degrades rapidly.

We next cxaminecl thc pcrformance of the receiver for fixed delay linc values and variable sampling instants. As shown in Appendix C, the $d\varphi/dt$ detector may be regarded as a limiting form of differential delay as the value of delay approaches zero. Fig. 19 shows the performance for the worst and best sequences as the sampling instant is varied across the bit. It is clear from this figure that the best sampling instant is in the middle of the bit, i.e., midway between transitions of the detector output wave. It turns out tbat tbe best sequence is tbe sequence of all marks or spaces, while the worst sequence is that of reversals.

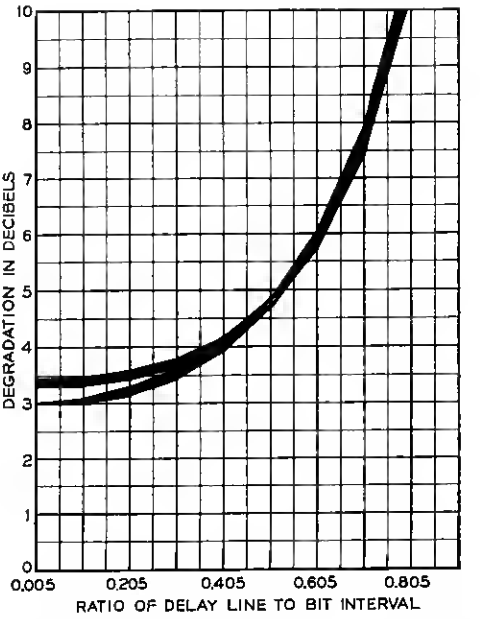

Fig. 13 — Db degradation vs differential delay with sampling at 0.005-bit interval.

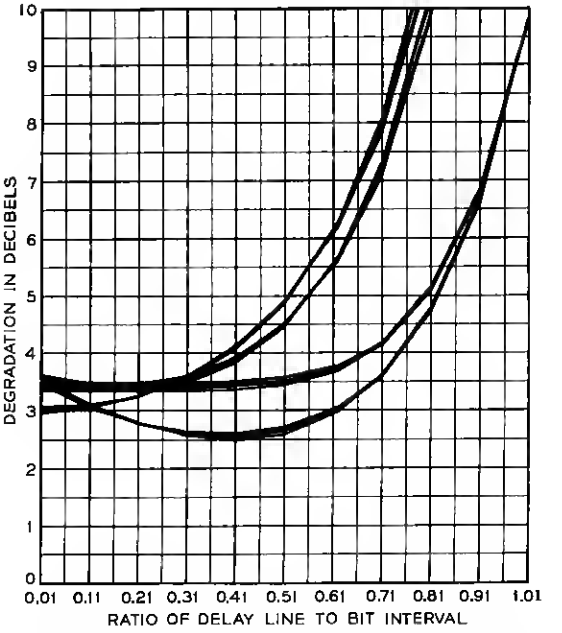

Fig. 14 — Db degradation vs differential delay with sampling at 0.11-bit interval.

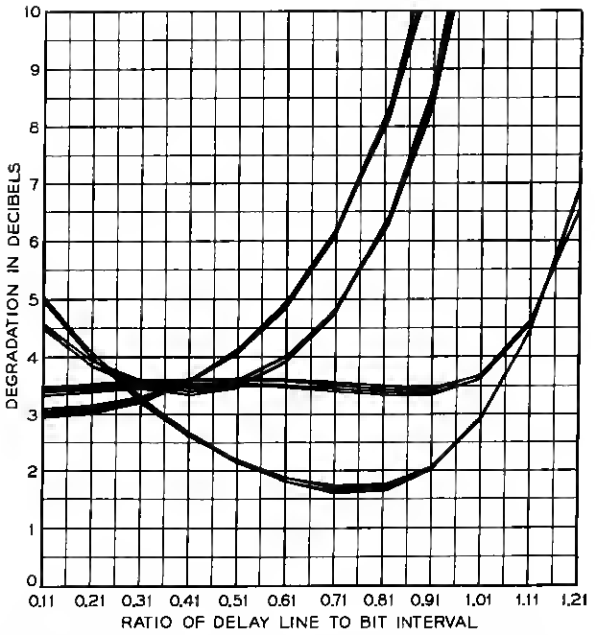

Fig. 15 — Db degradation vs differential delay with sampling at 0.21-bit interval.

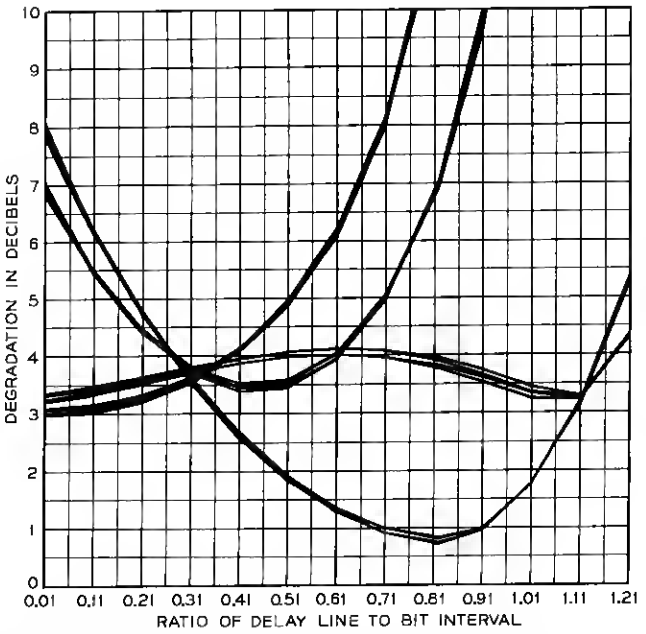

Fig. 16 — Db degradation vs differential delay with sampling at 0.31-bit interval.

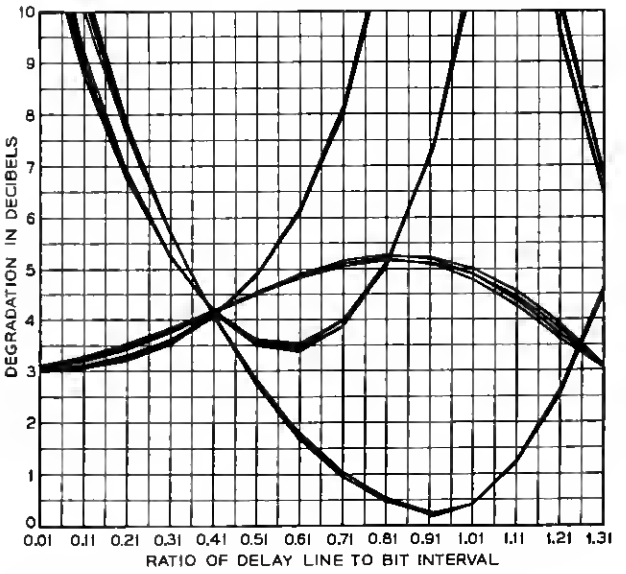

Fig. 17 — Db degradation vs differential delay with sampling at 0.41-bit interval.

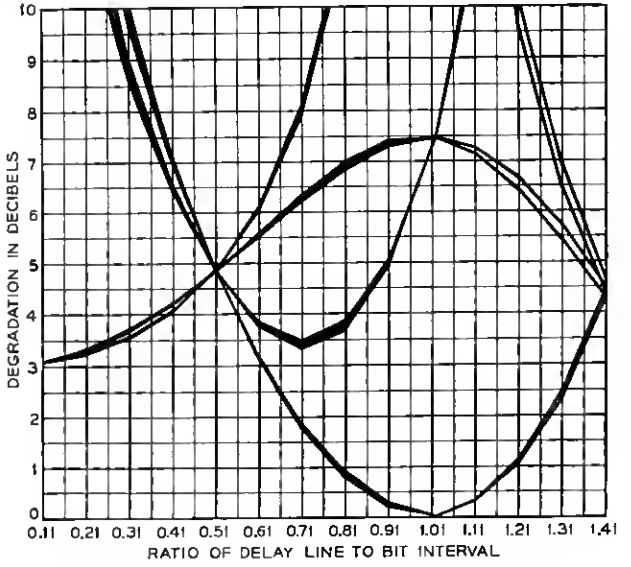

Fig. 18 — Db degradation vs differential delay with sampling at 0.51-bit interval.

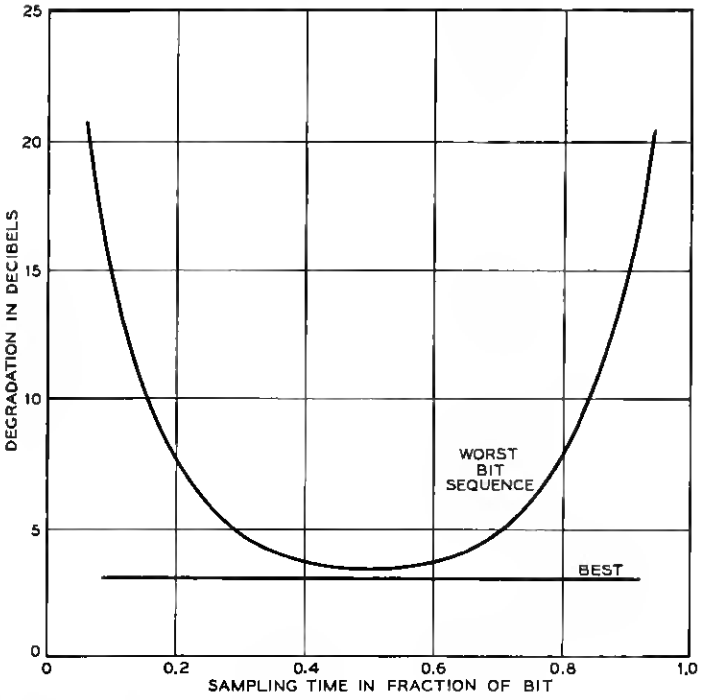

Fig. 19 — Db degradation vs sampling time for $d\varphi/dt$ detector with no delay distortion.

Fig. 20 exhibits the same information as Fig. 19 with the half-bit delay. In this case the worst sequence is a function of the sampling instant. At one-fourth bit and three-fourths bit sampling points referred to an origin at the transition points of the detector output, the intersymbol interference is zero, and at these instants the degradation is 4.8 db for all sequences. Nearer the center of the bit interval as seen from the detector output, the degradation is slightly greater. We nevertheless conclude that the best sampling instant is at this center, which we shall refer to as "mid-bit." By sampling at this time, we obtain a spread of half a bit interval tolerance to sampling jitter.

Similar curves are found when the receiver delay line is other than zero or a half-bit. In Fig. 21 we show the degradation of the worst sequence as a function of the receiver delay line. The zero and half-bit values are as shown on Figs. 19 and 20. The curve as shown applies to use of a delay line in which the phase shift is an odd multiple of $\pi/2$ at midband. We have previously shown that the signal can be recovered

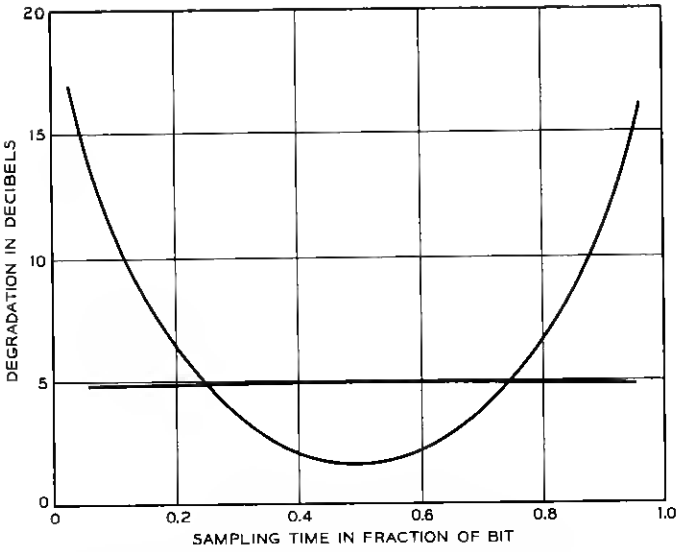

Fig. 20 — Db degradation vs sampling time for half-bit differential-delay detector with no delay distortion.

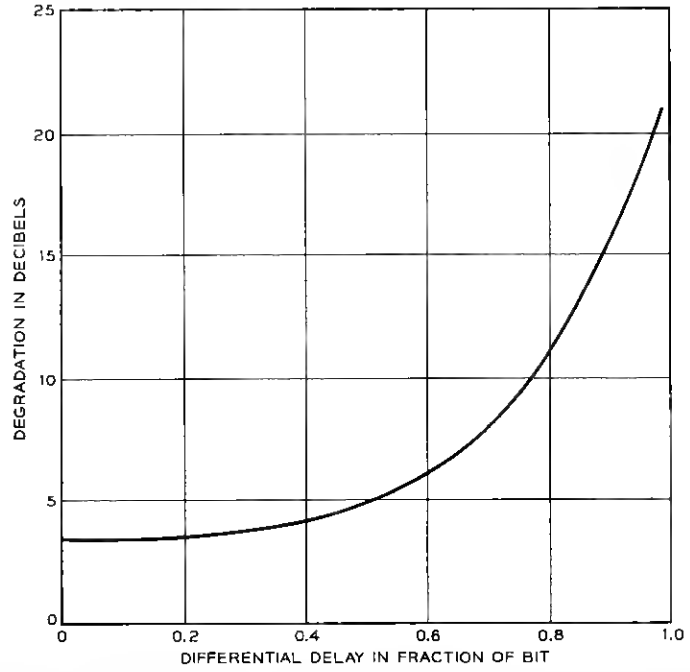

Fig. 21 — Db degradation vs differential delay for most vulnerable sequence sampled at mid-bit.

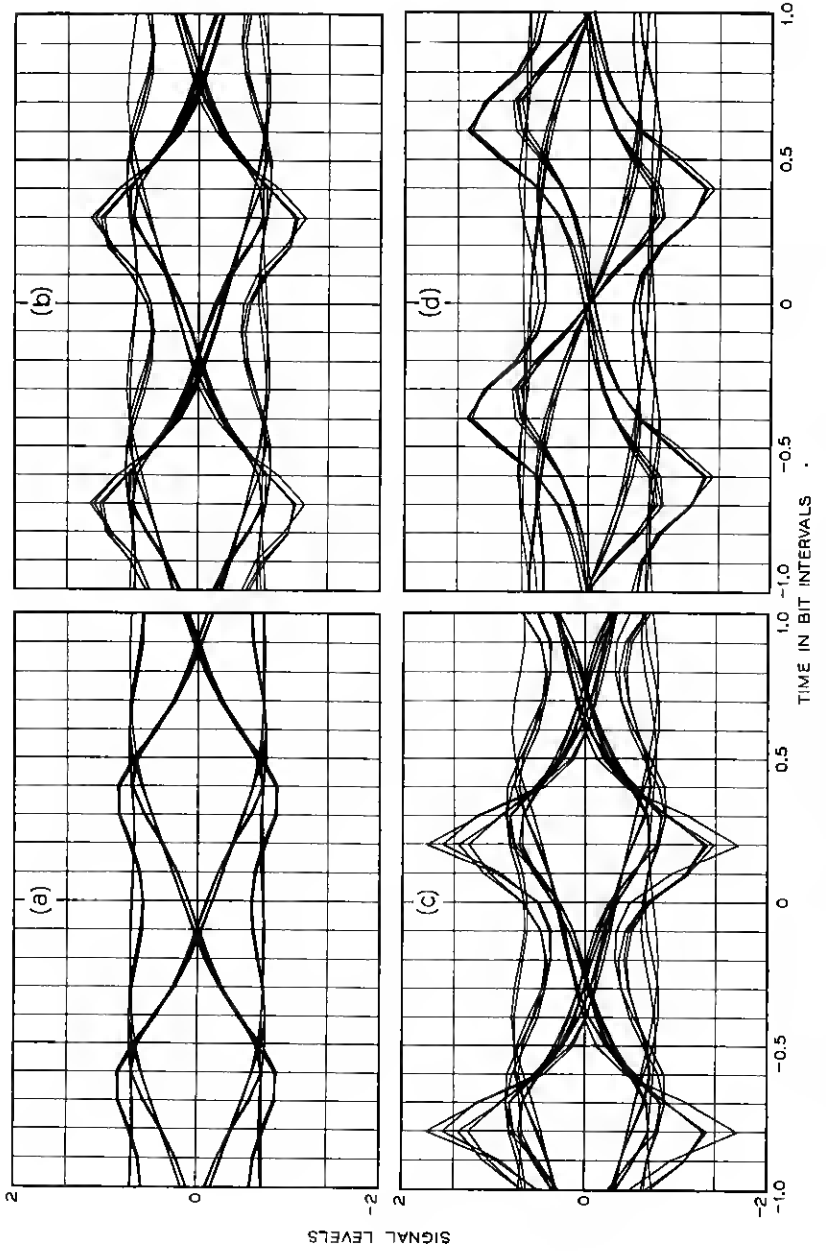

Fig. 22 — Eye patterns for $d\varphi/dt$ detector with delay distortion: (a) quadratic distortion of one bit interval; (b) quadratic, two bit intervals; (c) quadratic, three bit intervals; (d) linear, one bit interval.

with low degradation at a full-bit delay by differential encoding and use of a delay line with phase shift equal to a multiple of π at midband.

The above results indicate that, neglecting sampling jitter, least degradation results with least receiver delay, and the $d\varphi/dt$ receiver is best. Differences are small up to a half-bit. Addition of delay distortion alters this conclusion.

IX. EFFECTS OF DELAY DISTORTION

We preface our discussion by exhibiting sets of computed eye patterns for cases of linear and quadratic delay distortion in the channel. The amount of distortion is specified by the increment in delay measured in bit intervals between the center and edge of the transmission band. Results for the $d\varphi/dt$ detector are shown in Fig. 22 and for the half-bit delay differential detector in Fig. 23. The eye is found to close for smaller amounts of linear delay variation than for quadratic. As before, these results are not to be taken as quantitative measures of performance. The same numbers which determine the eye traces are also used in calculating error probabilities, but in a different way which precludes derivation of either final result directly from the other.

Fig. 24 shows the calculated degradation from the ideal for the $d\varphi/dt$ detector with various amounts of quadratic delay distortion measured in bit intervals. The results are plotted as a function of sampling time for the most vulnerable data sequence. There is an indication that the best sampling time is not at mid-bit when the delay distortion is large. However, the decreased tolerance to timing jitter would make such a shift undesirable. The effect of linear delay distortion, as exhibited in Fig. 25, is considerably worse. A comparison between effects of the two kinds of distortion is obtained in Fig. 26 by replotting the mid-bit sampling results of Figs. 24 and 25 as a function of delay distortion. As stated before, the $d\varphi/dt$ detector is equivalent to zero differential delay.

Figs. 27 and 28 present corresponding curves for the case of a half-bit differential delay. The shapes are similar to those for $d\varphi/dt$. There is slightly more tolerance to jitter, and the best sampling time is still at mid-bit.

In an effort to compare various receiver delay lines, previous data are cross-plotted in two ways. In Fig. 29, we in effect extend Fig. 21 to show how performance is affected by amount of differential delay when various amounts of specified linear and quadratic delay distortion are present in the channel. The curve for zero delay distortion replotted

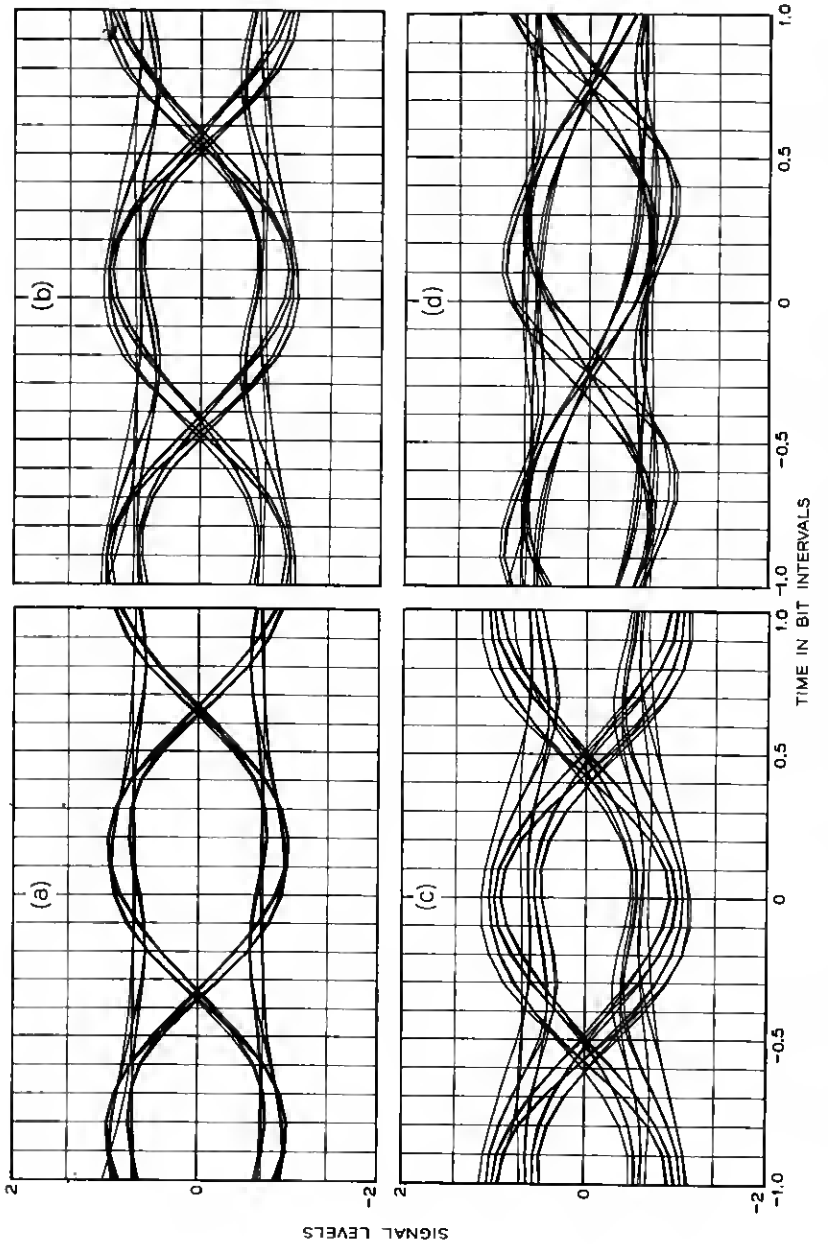

Fig. 23 — Eye patterns for half-bit differential delay detector with delay distortion: (a) quadratic distortion of one bit interval; (b) quadratic, two bit intervals; (c) quadratic, three bit intervals; (d) linear, one bit interval.

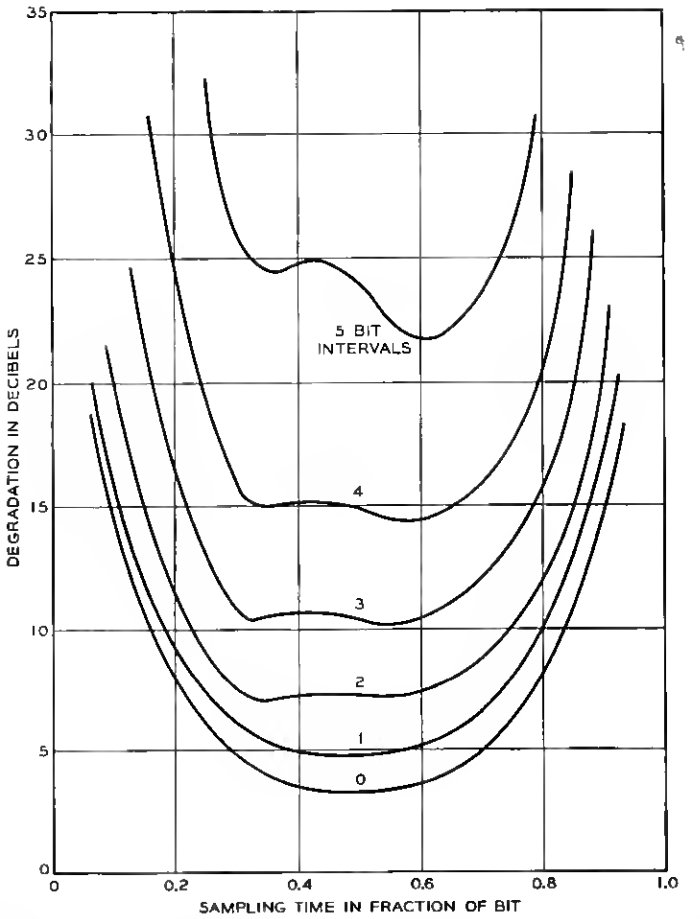

Fig. 24 — Db degradation of $d\varphi/dt$ detector vs sampling time for quadratic delay distortion.

from Fig. 21 shows the $d\varphi/dt$ receiver as best for a constant-delay channel. For quadratic delay distortion of two bit intervals or linear delay distortion of one bit interval, there is slightly less degradation at a half-bit of differential delay. For larger amounts of delay distortion the advantage from more receiver delay is greater.

The same conclusions can be drawn from Fig. 30, in which the abscissas are amounts of quadratic delay distortion and the curves are drawn for specified values of differential delay. The curves cross over from the condition of a preference for least differential delay with low delay

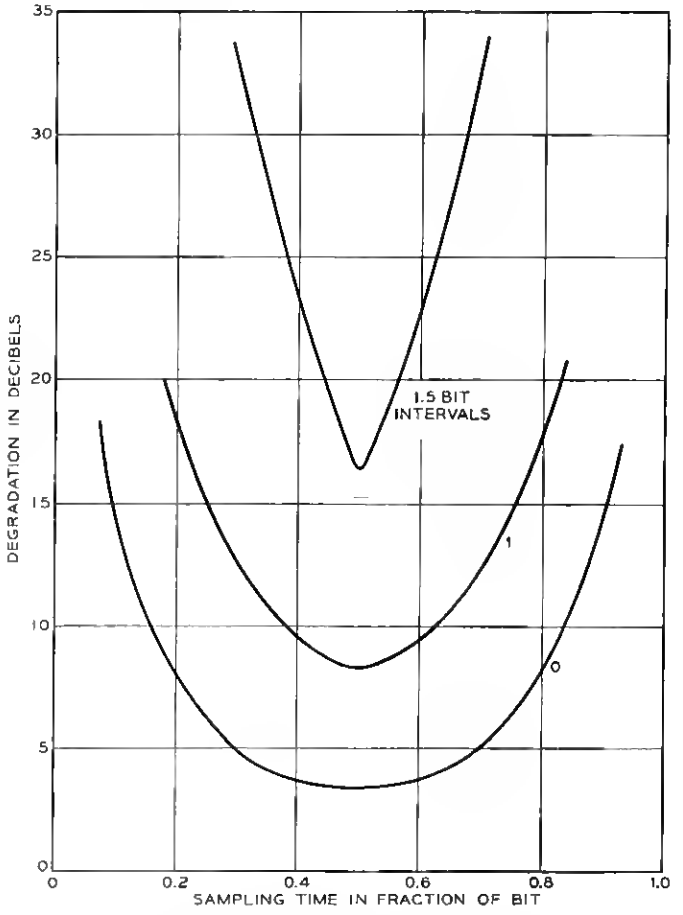

Fig. 25 — Db degradation of $d\varphi/dt$ detector vs sampling time for linear delay distortion.

distortion to a preference for greatest delay with high delay distortion. The exact choice of delay line depends on several factors. Say the maximum delay distortion is not greater than three hit intervals, then the half-hit delay is best over most of the range and only slightly worse over the rest. If the delay distortion is never more than one bit interval, a $d\varphi/dt$ detector is hest, while if very high delay distortions are encountered, and fairly high degradation is permissible at lower values, a delay line of 0.7 bit interval would be better. It was also found that the longer delay lines provide more tolerance to jitter in the sampling time.

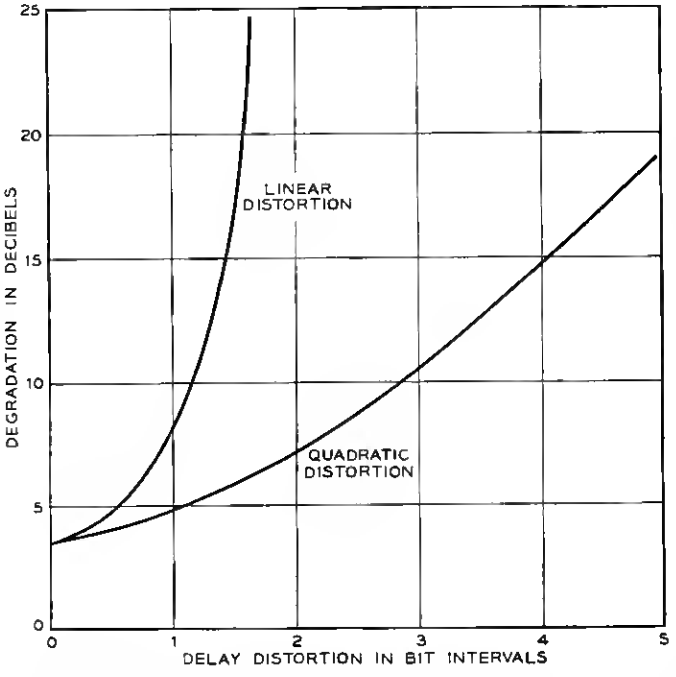

Fig. 26 — Db degradation of $d\varphi/dt$ detector vs linear and quadratic delay distortion.

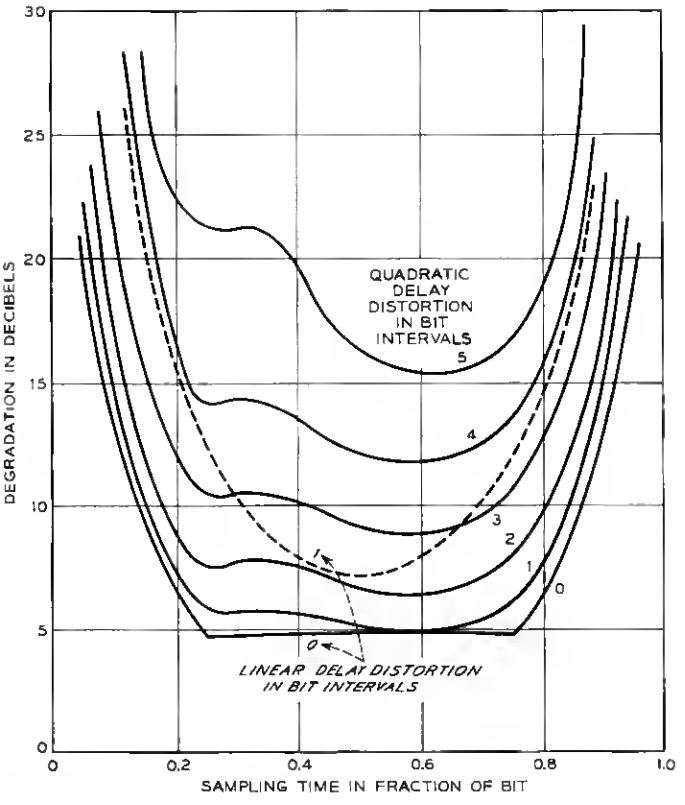

Fig. 27 — Db degradation with half-bit differential delay vs sampling time for quadratic delay distortion.

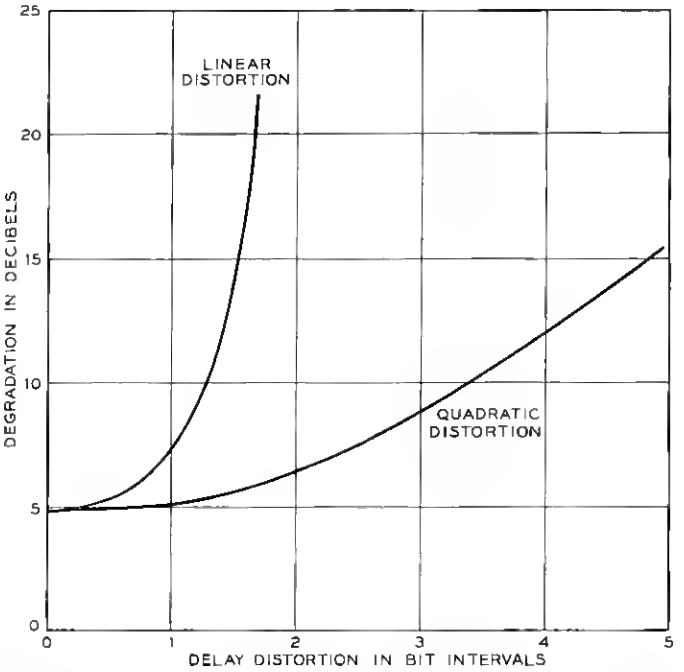

Fig. 28 — Db degradation with half-bit differential delay vs linear and quadratic delay distortion.

The asymptotic values of db impairment at large signal-to-noise ratio have been used throughout our discussion of Figs. 24–30 as a measure of performance for conditions of practical interest. As a check on the validity of this concept, complete curves of error probability vs signal-to-noise ratio have been computed from the exact formulas in representative cases. These curves are shown in Fig. 31 together with the asymptotic approximations. It appears that the latter are sufficient for most engineering applications.

APPENDIX A

Assume that a limiter is inserted in the undelayed input to the multiplier as shown in Fig. 6(h). The input to the limiter is then given by (13), which can also be written in the equivalent form

$$\begin{aligned} E(t) &= R(t) \cos [\omega_c t + \varphi(t)] \\ R(t) &= [x_1^2(t) + y_1^2(t)]^{\frac{1}{2}} \geqq 0 \\ \tan \varphi(t) &= y_1(t)/x_1(t). \end{aligned} \tag{71}$$

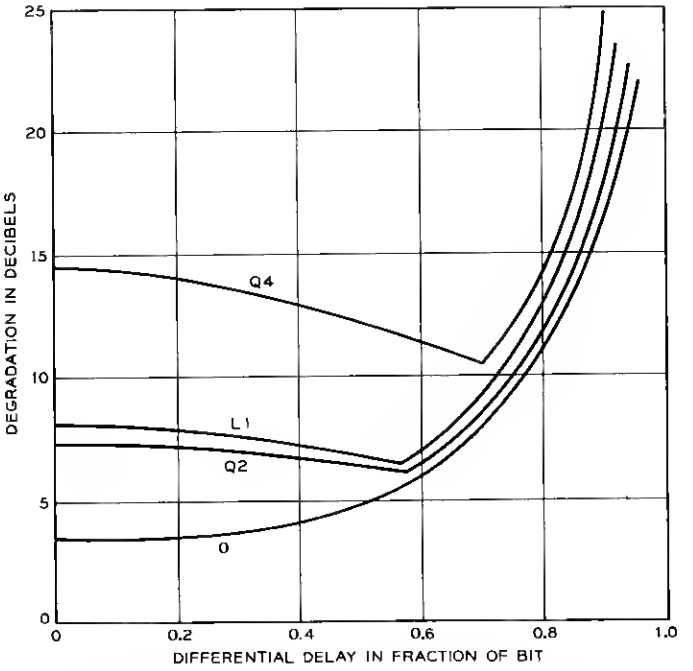

Fig. 29 — Db degradation vs differential delay for various amounts of linear and quadratic delay distortion. Curves are designated "L" for linear and "Q" for quadratic, followed by number of bit intervals of delay distortion.

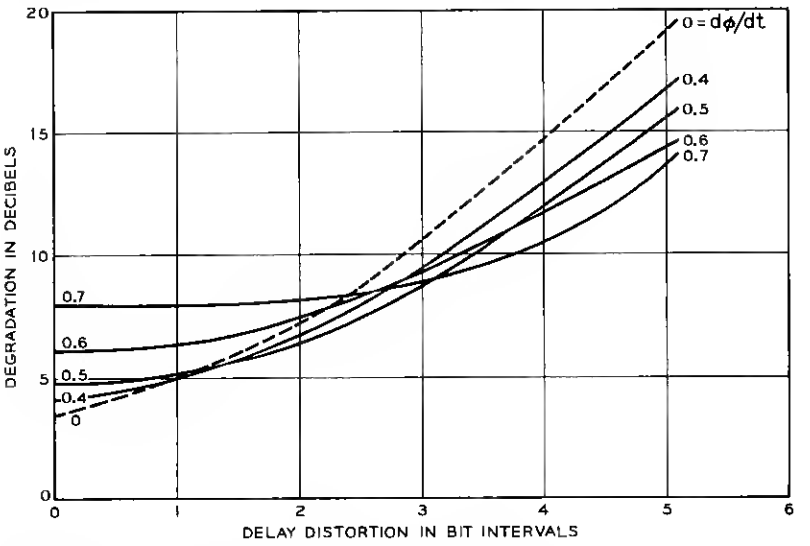

Fig. 30 — Db degradation vs quadratic delay distortion for various amounts of differential delay.

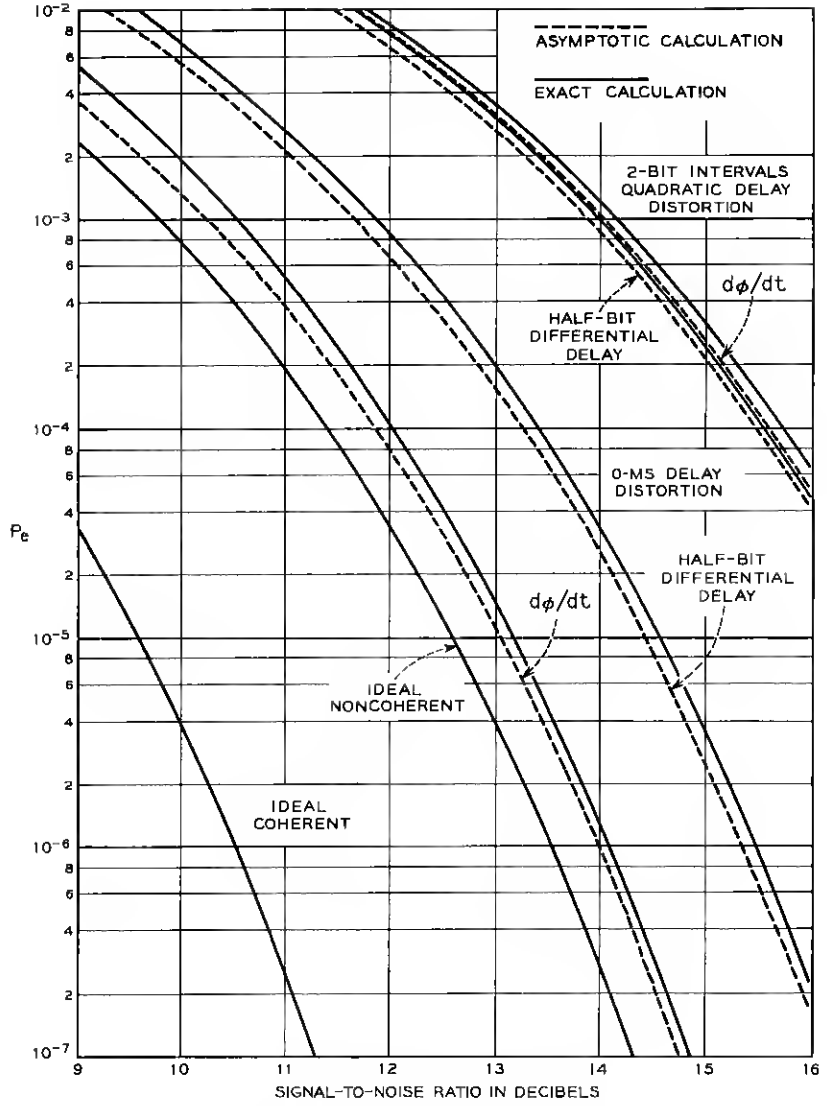

Fig. 31 — Comparison of results from exact and asymptotic formulas for error rates under representative conditions.

If the limiter is ideal, its output $E_L(t)$ is a positive constant E_0 when the value of the cosine function is positive and is a negative constant $-E_0$ when the cosine is negative. That is, if $z = \omega_c t + \varphi(t)$,

$$\begin{aligned} E_L(t) &= \begin{cases} E_0, & \cos z > 0 \\ -E_0, & \cos z < 0 \end{cases} \\ &= \frac{4E_0}{\pi} \sum_{n=0}^{\infty} \frac{(-)^n \cos (2n+1)z}{2n+1}. \end{aligned} \tag{72}$$

Multiplying $E_L(t)$ by the delayed signal and substituting $\cos \omega_c \tau = 0$, $\sin \omega_c \tau = -1$, gives:

$$\begin{aligned} E_L(t)E(t-\tau) = \frac{2E_0R(t-\tau)}{\pi} \sum_{n=0}^{\infty} \frac{(-)^n}{2n+1} \\ \cdot \{\sin [2n\omega_c t + (2n+1)\varphi(t) - \varphi(t-\tau)] \\ - \sin [2(n+1)\omega_c t + (2n+1)\varphi(t) \\ + \varphi(t-\tau)]\}. \end{aligned} \tag{73}$$

If it is possible by low-pass filtering to accept the term

$$\sin [\varphi(t) - \varphi(t-\tau)]$$

while rejecting the next higher-frequency terms $\sin [2\omega_c t + 3\varphi(t) - \varphi(t-\tau)]$ and $\sin [2\omega_c t + \varphi(t) + \varphi(t-\tau)]$, the filtered response becomes:

$$\begin{aligned} V_{lf}(t) &= \frac{2E_0R(t-\tau)}{\pi} \sin [\varphi(t) - \varphi(t-\tau)] \\ &= \frac{2E_0R(t-\tau)}{\pi} (\sin \varphi \cos \varphi_d - \cos \varphi \sin \varphi_d) \\ &= \frac{2E_0R(t-\tau)(y_1x_{1d} - x_1y_{1d})}{\pi(x_1^2 + y_1^2)^{\frac{1}{2}}(x_{1d}^2 + y_{1d}^2)^{\frac{1}{2}}}. \end{aligned} \tag{74}$$

The only term in (74) which can have both plus and minus signs is $y_1x_{1d} - x_1y_{1d}$, which is the same term found to be the basis of binary decisions in (17) for the case in which a pure product was taken. The switched modulator therefore gives the same error performance as the productor if there is sufficient frequency separation between the desired low-frequency output and the sidebands on $2\omega_c$.

An ideal limiter was assumed for simplicity in the argument just given, but the equivalence can be proved for a wide class of nonlinear devices in one of the two paths. It is sufficient that the output of the

device is of the same polarity as the input and can be expanded as a Fourier series in terms of the input values. We can then write

$$E_L(t) = F(R \cos z) = \frac{F_0}{2} + \sum_{n=1}^{\infty} F_n \cos nz \tag{75}$$

$$F_n = \frac{1}{\pi} \int_{-\pi/2}^{3\pi/2} F(R \cos z) \cos nz \, dz. \tag{76}$$

Note that R varies with time but cannot be negative.

In particular

$$F_1 = \frac{1}{\pi} \int_{-\pi/2}^{\pi/2} F(R \cos z) \cos z \, dz + \frac{1}{\pi} \int_{\pi/2}^{3\pi/2} F(R \cos z) \cos z \, dz. \tag{77}$$

In the first integral cos z is positive and hence $F(R \cos z)$ is positive. In the second integral cos z is negative, $F(R \cos z)$ is negative, and the product is positive. It follows that F_1 is positive.

If we repeat the calculation leading to (74) with the Fourier series (75) replacing (72), we find that $2E_0/\pi$ is replaced by F_1, and since F_1 is positive, the conclusion is the same. It will however be more difficult to isolate the desired low-frequency term when F_0 and F_2 are present, since these coefficients lead to components centered about ω_c. It is therefore preferable that the nonlinear function F have odd symmetry about the origin, in which case the even-order coefficients vanish.

The more general argument is useful in showing that the switched modulator does not have to be perfect. Note that the equivalence is destroyed if limiters are inserted in both paths. There would then be sidebands on harmonics in both inputs to the modulator, and the beats between these sidebands would generate additional components in the low-frequency band.

APPENDIX B

Simplification of the Error Probability Formula and Determination of Its Asymptotic Form

First, replace the parameter b by

$$k = \frac{b}{\rho} = \frac{\bar{z}_2\bar{z}_4 - \bar{z}_1\bar{z}_3}{2\sigma_1\sigma_2\rho^2}. \tag{78}$$

The expression to be evaluated is now

$$\Lambda_1(\rho,a,k) = \frac{1}{2} - \frac{\rho}{2\sqrt{\pi}} \int_{-1}^{1} e^{-\rho^2 x^2} \operatorname{erf} [\rho(a\sqrt{1-x^2} - kx)]dx. \quad (79)$$

Substitute $\rho x = x'$ and replace the error function by its definition as an integral. The result is

$$\Lambda_1(\rho,a,k) = \frac{1}{2} - \frac{1}{\pi} \int_{-\rho}^{\rho} \int_{0}^{a\sqrt{\rho^2-x^2}-kx} e^{-(x^2+y^2)}dx\,dy. \quad (80)$$

A graph of the region of integration in the xy-plane is shown in Fig. 32. From this figure, we deduce that by a change to polar coordinates, we obtain the equivalent expression

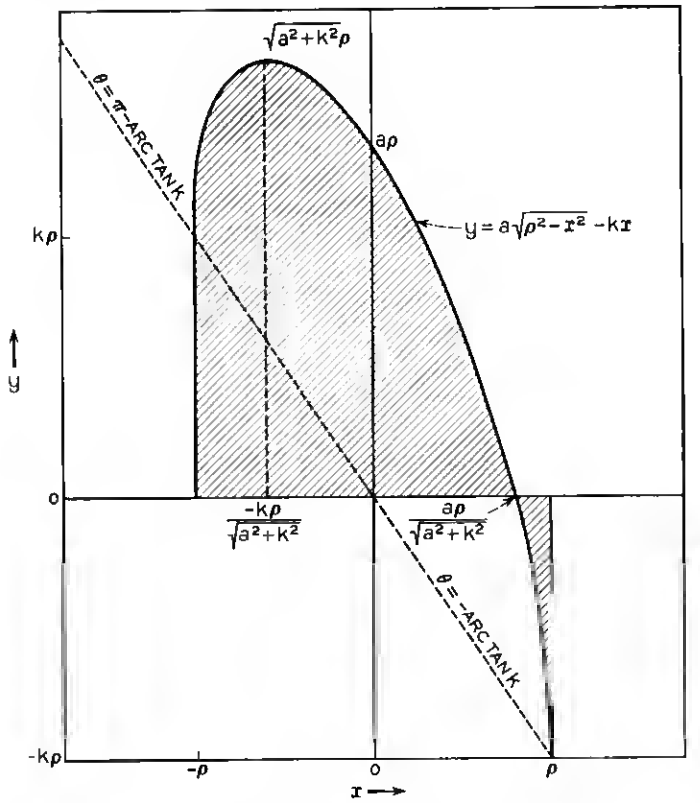

Fig. 32 — Region of integration for evaluation of double integral in (80).

$$\begin{aligned}\Lambda_1(\rho,a,k) = \frac{1}{2} &- \frac{1}{\pi}\int_0^{\pi-\arctan k} d\theta \int_0^{f_1(\theta)} e^{-r^2} r\, dr \\ &- \frac{1}{\pi}\int_{\pi-\arctan k}^{\pi} d\theta \int_0^{-\rho/\cos\theta} e^{-r^2} r\, dr \\ &- \frac{1}{\pi}\int_0^{-\arctan k} d\theta \int_{f_1(\theta)}^{\rho/\cos\theta} e^{-r^2} r\, dr\end{aligned} \tag{81}$$

where

$$f_1(\theta) = A\rho[(\sin\theta + k\cos\theta)^2 + a^2\cos^2\theta]^{-1/2}. \tag{82}$$

The integration with respect to r can be performed, and after some subsequent combining of terms, we find

$$\Lambda_1(\rho,a,k) = \frac{1}{2\pi}\int_0^{\pi} \exp\left[-\frac{a^2\rho^2}{(\sin\theta + k\cos\theta)^2 + a^2\cos^2\theta}\right] d\theta. \tag{83}$$

By the substitution $2\theta = \varphi$, (83) transforms to

$$\begin{aligned}\Lambda_1(\rho,a,k) &= \frac{1}{4\pi}\int_0^{2\pi} \\ &\cdot \exp\left[-\frac{2a^2\rho^2}{k^2 + a^2 + 1 + (k^2 + a^2 - 1)\cos\varphi - 2k\sin\varphi}\right] d\varphi.\end{aligned} \tag{84}$$

We then note that

$$\begin{aligned}&(k^2 + a^2 - 1)\cos\varphi - 2k\sin\varphi \\ &\quad = [(k^2 + a^2 - 1)^2 + 4k^2]^{\frac{1}{2}} \cos\left(\varphi + \arctan\frac{2k}{k^2 + a^2 - 1}\right).\end{aligned} \tag{85}$$

Taking advantage of the fact that the range of integration is one complete period of the integrand in φ, we can replace the sum of φ and a constant angle by a new variable without changing the limits. Noting furthermore that the integrand then becomes an even function of the variable of integration, we obtain finally

$$\begin{aligned}\Lambda_1(\rho,a,k) &= G(c,d) \\ &= \frac{1}{2\pi}\int_0^{\pi} \exp\left[-\frac{c^2}{1 + d^2\cos\varphi}\right] d\varphi\end{aligned} \tag{86}$$

where c^2 and d^2 have the values given by (30) and (31).

Applying the method of steepest descents to the case in which c^2 is large, we write

$$G(c,d) = \frac{1}{2\pi}\int_0^{\pi} e^{-c^2\Phi(\varphi)}d\varphi \tag{87}$$

$$\begin{aligned}
\Phi(\varphi) &= (1 + d^2\cos\varphi)^{-1} \\
\Phi'(\varphi) &= d^2(1 + d^2\cos\varphi)^{-2}\sin\varphi \\
\Phi'(\varphi_0) &= 0 \text{ at } \varphi_0 = 0 \text{ or } \pi \\
\Phi''(\varphi_0) &= d^2(1 + d^2\cos\varphi_0)^{-2}\cos\varphi_0 > 0 \quad \text{for} \quad \varphi_0 = 0.
\end{aligned} \tag{88}$$

Hence, we set $\varphi_0 = 0$ and approximate $\Phi(\varphi)$ by

$$\begin{aligned}
\Phi(\varphi) &= \Phi(\varphi_0) + \tfrac{1}{2}\Phi''(\varphi_0)(\varphi - \varphi_0)^2 + \cdots \\
&= (1 + d^2)^{-1} + \tfrac{1}{2}d^2(1 + d^2)^{-2}\varphi^2 + \cdots .
\end{aligned} \tag{89}$$

If $c^2 \gg 1$,

$$\begin{aligned}
G(c,d) &\sim \frac{1}{2\pi}\exp\left(-\frac{c^2}{1 + d^2}\right)\int_0^{\pi}\exp\left[-\frac{c^2d^2\varphi^2}{2(1 + d^2)^2}\right]d\varphi \\
&\sim \frac{1 + d^2}{2cd\sqrt{2\pi}}\exp\left(-\frac{c^2}{1 + d^2}\right).
\end{aligned} \tag{90}$$

APPENDIX C

Limit of Performance as Delay Approaches Zero

As the delay τ is made small, the delayed variables in (17) can be expressed in terms of the undelayed ones by the linear approximations,

$$\begin{aligned}
x_{1d} &\approx x_1 - \tau\dot{x}_1 \\
y_{1d} &\approx y_1 - \tau\dot{y}_1 ,
\end{aligned} \tag{91}$$

where the dot signifies the derivative with respect to time. In the limit we then find

$$2E_{1f} \approx \tau(x_1\dot{y}_1 - y_1\dot{x}_1) . \tag{92}$$

The amplitude of the detected signal approaches zero as the delay is made small, but if τ is not actually zero, the lack of output can be compensated by linear amplification. Hence in the absence of imperfections other than additive noise in the channel, binary decisions are made on the basis of the sign of $x\dot{y}_1 - y_1\dot{x}_1$.

In the case of a $d\varphi/dt$ detector, binary decisions are made on the basis of the sign of

$$\frac{d\varphi}{dt} = \frac{d}{dt} \arctan \frac{y_1}{x_1} = \frac{x_1\dot{y}_1 - y_1\dot{x}_1}{x_1^2 + y_1^2}. \tag{93}$$

Since $x_1^2 + y_1^2$ cannot change sign, the decisions are actually made on the sign of $x_1\dot{y}_1 - y_1\dot{x}_1$. We conclude that the binary error rates of the two systems must approach equality as τ approaches zero.

APPENDIX D

Computational Details

For the pulse response (36) we use the impulse response corresponding to a raised-cosine spectrum, namely

$$g(t) = \frac{\sin\left(\frac{2\pi}{T}t\right)}{\pi\left(\frac{2t}{T}\right)\left[1 - \left(\frac{2t}{T}\right)^2\right]}. \tag{94}$$

The sampled value of the pulse train at $t = T + \epsilon T$ is then given by

$$s_1[T(1+\epsilon)] = \sum_n (-1)^n b_n \frac{\sin 2\pi[1 - \delta_2 - n]}{2\pi(1 - \delta_2 - n)[1 - 4(1 - \delta_2 - n)^2]} \equiv s_1^0(\delta_2) \tag{95}$$

and the delayed version after τ seconds by

$$s_1[T(1 + \epsilon - \delta)] = \sum_n (-1)^n b_n \frac{\sin 2\pi[1 - \delta_1 - n]}{2\pi(1 - \delta_1 - n)[1 - 4(1 - \delta_1 - n)^2]} = s_1^0(\delta_1) \tag{96}$$

where

$$\delta = \tau/T, \qquad \delta - \epsilon = \delta_1, \qquad \delta_1 - \delta = \delta_2. \tag{97}$$

In terms of the two new variables δ_1, δ_2, the noise correlation is given by

$$r(\delta) = \frac{\sin 2\pi(\delta_1 - \delta_2)}{2\pi(\delta_1 - \delta_2)[1 - 4(\delta_1 - \delta_2)^2]}. \tag{98}$$

The observed voltage or current ξ can now be written as a function of δ_1 and δ_2 in the following form:

$$\xi(\delta_1, \delta_2) = [As_1^0(\delta_2) + x][-A \sin \pi\delta_1 - y_d] + [A \sin \pi\delta_2 + y][As_1^0(\delta_1) + x_d]. \tag{99}$$

In the calculation of eye patterns, the noise samples x, x_d, y, and y_d are omitted. By dropping these terms in (99) and substituting successively $\epsilon = 0$ and $\epsilon = \delta$, we verify that the samples at these instants depend on only one value in the data sequence. This confirms the statements made in Section III in the discussion of eye patterns. When the noise samples are inserted, the entire previous signal history exerts its effect at all sampling instants, and in fact the preferred sampling instant from the standpoint of low error rate and tolerance to jitter in sampling time is not necessarily the one in which intersymbol interference vanishes in the absence of noise.

The probability of error for a particular sequence is:

$$\begin{aligned} P_e = {} & \tfrac{1}{2} \Pr \{[As_1^{+}(\delta_2) + x][-A \sin \pi\delta_1 - y_d] \\ & + [A \sin \pi\delta_2 + y][As_1^{+}(\delta_1) + x_d] < 0\} \\ & + \tfrac{1}{2} \Pr \{[As_1^{-}(\delta_2) + x][-A \sin \pi\delta_1 - y_d] \\ & + [A \sin \pi\delta_2 + y][As_1^{-}(\delta_1) + x_d] \geqq 0\} \end{aligned} \tag{100}$$

where

$$\begin{aligned} & s_1^{\pm}(u) \\ & = \sum_{n \neq 1} (-1)^n b_n \frac{\sin 2\pi[1 - \mu - n]}{2\pi(1 - \mu - n)[1 - 4(1 - \mu - n)^2] \pm r(\mu)}. \end{aligned} \tag{101}$$

Applying the transformation in (21) we obtain:

$$\begin{aligned} \bar{z}_1^{\pm} &= (A/2)[s_1^{\pm}(\delta_2) - \sin \pi\delta_2 + s_1^{\pm}(\delta_1) - \sin \pi\delta_1] \\ \bar{z}_2^{\pm} &= (A/2)[s_1^{\pm}(\delta_2) + \sin \pi\delta_2 + s_1^{\pm}(\delta_1) + \sin \pi\delta_1] \\ \bar{z}_3^{\pm} &= (A/2)[-s_1^{\pm}(\delta_2) + \sin \pi\delta_2 + s_1^{\pm}(\delta_1) - \sin \pi\delta_1] \\ \bar{z}_4^{\pm} &= (A/2)[s_1^{\pm}(\delta_2) + \sin \pi\delta_2 - s_1^{\pm}(\delta_1) - \sin \pi\delta_1]. \end{aligned} \tag{102}$$

The required parameters in (25)–(27) are

$$\rho^2 = \frac{M}{8A^2} \frac{\bar{z}_1^{\,2} + \bar{z}_2^{\,2}}{1 + r(\delta_1 - \delta_2)} \tag{103}$$

$$a = \frac{\bar{z}_1\bar{z}_4 + \bar{z}_2\bar{z}_3}{\bar{z}_1^{\,2} + \bar{z}_2^{\,2}} \sqrt{\frac{1 + r(\delta_1 - \delta_2)}{1 - r(\delta_1 - \delta_2)}} \tag{104}$$

$$k = \frac{\bar{z}_2\bar{z}_4 - \bar{z}_1\bar{z}_3}{\bar{z}_1^{\,2} + \bar{z}_2^{\,2}} \sqrt{\frac{1 + r(\delta_1 - \delta_2)}{1 - r(\delta_1 - \delta_2)}}. \tag{105}$$

As pointed out before, the asymptotic degradation is given by

$$\text{degradation} = 10 \log_{10} (1/\kappa)\text{db} \tag{106}$$

where

$$\kappa = \frac{k^2 + a^2 + 1 + [(k^2 + a^2 - 1)^2 + 4k^2]^{\frac{1}{2}}}{2a^2\rho_1^2} \tag{107}$$

and

$$\rho = M\rho_1 . \tag{108}$$

When delay distortion is present in the channel, a convolution is performed to evaluate the resulting in-phase and quadrature components of the noise-free input to the detector. It is convenient to combine the assumed delay distortion with the equivalent amplitude characteristic arising from signal pulse shaping, sending filter, transmission line, and receiving filter to form a single complex transmittance function

$$H(\omega) = A(\omega)e^{j\theta(\omega)} . \tag{109}$$

The impulse response of the medium is then

$$\begin{aligned} h(t) &= \int_{-\infty}^{\infty} H(\omega)e^{j\omega t}\, d\omega/2\pi \\ &= \int_{-\infty}^{\infty} A(\omega)e^{j[\theta(\omega)+\omega t]}\, d\omega/2\pi \\ &= \int_{0}^{\infty} A(\omega)\cos\,[\theta(\omega) + \omega t]\, d\omega/\pi . \end{aligned} \tag{110}$$

Let $A(\omega) = B(\omega - \omega_c) = B(\upsilon)$, $\theta(\omega) = \varphi(\omega - \omega_c) + \varphi_0 = \varphi(\upsilon) + \varphi_0$, and decompose (110) into the in-phase and quadrature components

$$\begin{aligned} h(t) &= \int_{-\omega_0}^{\infty} B(\upsilon)\cos\,[\varphi(\upsilon) + \upsilon t + \omega_c t + \varphi_0]\, d\upsilon/\pi \\ &= h_1(t)\cos\,\omega_c t - h_2(t)\sin\,\omega_c t, \end{aligned} \tag{111}$$

where

$$\begin{aligned} h_1(t) &= \int_{-\omega_0}^{\infty} B(\upsilon)\cos\,[\varphi(\upsilon) + \upsilon t + \varphi_0]\, d\upsilon/\pi \\ h_2(t) &= \int_{-\omega_0}^{\infty} B(\upsilon)\sin\,[\varphi(\upsilon) + \upsilon t + \varphi_0]\, d\upsilon/\pi . \end{aligned} \tag{112}$$

Since the medium is assumed to be linear, the signal input to the detector is the convolution of the input and the impulse response,

$$V_1(t) = \int_{-\infty}^{\infty} V(t - \tau)h(\tau)d\tau , \tag{113}$$

where $V(t)$ can be written in the form:

$$V(t) = P_0(t) \cos \omega_c t - Q_0(t) \sin \omega_c t. \tag{114}$$

Then from (111),

$$V_1(t) = \int_{-\infty}^{\infty} [P_0(t - \tau) \cos (\omega_c t - \omega_c \tau) - Q_0(t - \tau) \sin (\omega_c t - \omega_c \tau)][h_1(\tau) \cos \omega_c \tau - h_2(\tau) \sin \omega_c \tau] d\tau . \tag{115}$$

Dropping the double-frequency terms, we obtain for the input to the detector

$$V_r(t) = P(t) \cos \omega_c t - Q(t) \sin \omega_c t \tag{116}$$

where

$$\begin{aligned} P(t) &= \int_{-\infty}^{\infty} [P_0(t - \tau)h_1(\tau) - Q_0(t - \tau)h_2(\tau)] d\tau/2 \\ Q(t) &= \int_{-\infty}^{\infty} [P_0(t - \tau)h_2(\tau) + Q_0(t - \tau)h_1(\tau)] d\tau/2 . \end{aligned} \tag{117}$$

REFERENCES

1. Davey, J. R., Signal Space Diagrams, B.S.T.J., **43**, Nov., 1964, pp. 2973–2983.
2. Brand, S., and Carter, C. W., A 1,650-Bit-Per-Second Data System for Use over the Switched Telephone Network, A.I.E.E. Trans., Pt. I, Comm. and Elect., **80**, Jan., 1962, pp. 652–661.
3. Bennett, W. R., and Salz, J., Binary Data Transmission by FM Over a Real Channel, B.S.T.J., **42**, Sept., 1963, pp. 2387–2426.
4. Sunde, E. D., Ideal Pulses Transmitted by AM and FM, B.S.T.J., **38**, Nov., 1959, pp. 1357–1426.
5. Nyquist, H., Certain Topics in Telegraph Transmission, Trans., A.I.E.E **47**, pp. 617–644, Apr., 1928.